

Learning Analytics: um mapeamento sistemático

Barbara Moissa

Programa de Pós Graduação em
Computação Aplicada (PPGCA)
Universidade do Estado de Santa
Catarina (UDESC)
Joinville, SC, Brasil
barbara.moissa@gmail.com

Isabela Gasparini

Universidade do Estado de Santa
Catarina (UDESC)
Joinville, SC, Brasil
isabela.gasparini@udesc.br

Avanilde Kemczinski

Programa de Pós Graduação em
Computação Aplicada (PPGCA)
Universidade do Estado de Santa
Catarina (UDESC)
Joinville, SC, Brasil
avanilde.kemczinski@udesc.br

ABSTRACT

Learning Analytics (LA) is a new research field, related to the collection and analysis of students' data to enhance learning and the environment on which it occurs. This paper presents a systematic mapping of the literature on the field, seeking to understand which are the data types, the techniques, the stakeholders and the interventions; and their relation with the analysis goals: Monitoring and Analysis; Prediction and Intervention; Tutoring and Mentoring; Assessment and Feedback; Adaptation; Personalization and Recommendation; and Reflection. We also analyze LA in the Massive Open Online Courses (MOOCs) context. The results show several types of data (the main is navigational data) and techniques; the main stakeholders as researchers; and little focus on interventions. It also shows research opportunities in the MOOCs context.

RESUMO

Learning Analytics é uma nova área de pesquisa, relacionada à coleta de dados dos alunos para melhorar o aprendizado e o ambiente em que este ocorre. Este artigo apresenta um mapeamento sistemático da literatura da área, procurando entender quais são os tipos de dados, as técnicas, as pessoas interessadas e as intervenções; e sua relação com os objetivos de análise: Monitoramento e Análise; Predição e Intervenção; Tutoria e Mentoria; Avaliação e *Feedback*; Adaptação; Personalização e Recomendação; e Reflexão. Nós também analisamos a área no contexto dos *Massive Open Online Courses* (MOOCs). Os resultados mostram diferentes tipos de dados (o principal são os dados navegacionais) e técnicas; as principais pessoas interessadas são pesquisadores; e pouco foco nas intervenções. Estes também mostram oportunidades de pesquisa no contexto dos MOOCs.

Descritor de Categorias e Assuntos

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Termos Gerais

Human Factors

Palavras Chaves

Learning Analytics, análise de dados educacionais, *Massive Open Online Courses*, Informática na Educação

1. INTRODUÇÃO

Nos Ambientes Virtuais de Aprendizagem (AVAs) uma grande quantidade de dados de navegação pode ser coletada, possibilitando que sejam feitas diversas análises relacionadas ao comportamento do aluno no ambiente, auxiliando na prática pedagógica e até mesmo na melhoria do ambiente (de forma similar ao uso de técnicas *web analytics*, adotadas no comércio eletrônico). Neste sentido, surgem as *Learning Analytics* (LA), uma área de pesquisa relativamente nova na qual são analisados os dados produzidos por alunos dentro e fora dos AVAs com o objetivo de melhorar o processo de ensino-aprendizagem.

Com o objetivo de entender melhor a área, este artigo propõe um mapeamento sistemático da literatura para identificar quais os tipos de dados, quais técnicas, quais as intervenções (ações realizadas após uma análise para mudar o processo de ensino-aprendizagem) e quais os usuários aos quais os resultados são destinados, bem como qual a relação destes elementos com os objetivos de análise. Em seguida, para identificar oportunidades de estudo da área com dados de *Massive Open Online Courses* (MOOCs) fazemos uma análise dos trabalhos inseridos neste contexto.

Para possibilitar um melhor entendimento das questões de pesquisa e dos resultados, a Seção 2 contextualiza a área de *Learning Analytics*. Em seguida, a Seção 3 descreve a metodologia adotada para realizar o mapeamento sistemático proposto. A Seção 4 mostra os resultados gerais obtidos e a Seção 5 mostra os resultados no contexto dos MOOCs. A Seção 6 discute os resultados e, por fim, a Seção 7 apresenta as considerações finais, seguida das referências bibliográficas.

2. LEARNING ANALYTICS

Learning Analytics são a medida, coleta, análise e relato dos dados de estudantes e seus contextos com o propósito de entender e otimizar o aprendizado e o ambiente em que este ocorre [1]. Aplicado a disciplinas, cursos e departamentos, este processo fornece informações valiosas sobre o que está realmente acontecendo e sugere maneiras através das quais os educadores podem melhorar o processo de ensino-aprendizagem, além de indicar quais alunos podem desistir do curso ou aqueles que precisam de ajuda adicional para melhorar seu desempenho [2].

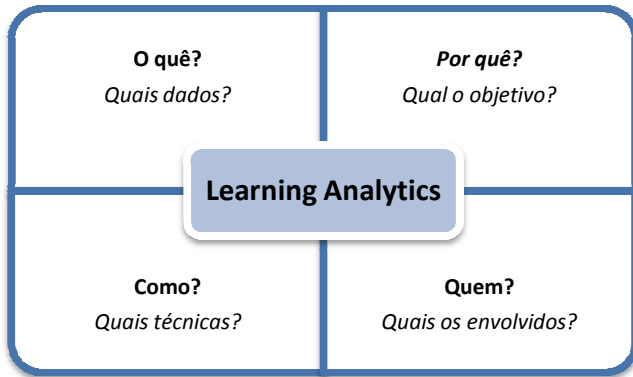


Figura 1. Modelo de referência de Learning Analytics [3]

O modelo de referência ilustrado na Figura 1, foi proposto com base na evolução da área para possibilitar a classificação da literatura e também para facilitar o seu entendimento [3]. Este modelo possui as mesmas dimensões adotadas por Atif *et al.* [4] e é baseado em quatro dimensões:

- **O quê?** Esta dimensão refere-se aos tipos de dados de dados coletados. Esses dados podem ser provenientes do ambiente de aprendizagem utilizado, de fontes institucionais, de redes sociais como o Facebook e o Twitter, etc.
- **Quem?** Esta dimensão refere-se a quem os resultados da análise serão direcionados, podendo ser a alunos, professores, instituições educacionais (administradores), pesquisadores, projetistas de sistemas, etc. Cada um dos envolvidos tem suas próprias perspectivas, objetivos e expectativas em relação aos resultados.
- **Por quê?** Esta dimensão está relacionada com os resultados/objetivos da análise realizada. Conforme Chatti *et al.* [3] e Atif *et al.* [4] estes objetivos são:
 - Monitoramento e Análise: como o nome diz tem como objetivo monitorar as ações do aluno e gerar relatórios com os resultados das análises.
 - Predição e Intervenção: desenvolver um modelo que tente prever o desempenho do aluno no futuro com base em suas ações e realizações atuais, fornecendo indicativos de alunos que precisam de atenção especial para então intervir.
 - Tutoria e Mentoria: ajudar o aluno em módulos específicos do curso (tutoria) ou durante todo o curso (mentoria).
 - Avaliação e *Feedback*: procura auxiliar na avaliação do aprendizado fornecendo o *feedback* adequado.
 - Adaptação: busca dizer ao aluno o que fazer através da reorganização dos recursos de aprendizagem e das atividades de acordo com as suas necessidades.
 - Personalização e Recomendação: busca permitir que o aluno defina como vai aprender (personalização) — enfatizando a busca pelo conhecimento —, mas fornecendo mecanismos (recomendação) para auxiliá-lo nesta busca de modo a evitar a sobrecarga de informações.
 - Reflexão: permite que as informações sejam comparadas entre alunos, turmas e cursos permitindo a verificação da eficácia das práticas de ensino e/ou aprendizagem.

- **Como?** Esta dimensão refere-se às diferentes técnicas que podem ser empregadas para detectar padrões contidos nos dados e atingir os objetivos contidos na dimensão *Por quê?*. De acordo com Chatti *et al.* [3] a área emprega uma grande variedade de métodos de outras áreas como, por exemplo:
 - *Academic Analytics*: uma categoria ampla de aplicações e tecnologias para coletar, armazenar, analisar e fornecer acesso a dados para ajudar usuários de companhias a tomar decisões melhores [5]. Na literatura, os trabalhos desta área geralmente são relativos a identificação de estudantes em risco (predição) [3]. Estas análises são aplicadas a nível institucional, regional (i.e., cidade, estado) e nacional/internacional [2].
 - Mineração de Dados Educacionais (do inglês *Educational Data Mining — EDM*): é uma área emergente, cujo objetivo é definir métodos para explorar dados provenientes de ambientes educacionais e utilizá-los para compreender melhor os alunos e seu contexto de aprendizagem [6]. Por mais que a definição desta área seja similar a definição de LA, ambas empregam técnicas diferentes. Enquanto a Mineração de Dados Educacionais emprega técnicas comuns de mineração de dados (e.g, clusterização, classificação, mineração de regras de associação) as LA podem empregar, além destas, outras técnicas (e.g., Análise das Redes Sociais, Estatísticas, Visualização da Informação). Outras diferenças entre as áreas podem ser encontradas em [7].

Cada uma das dimensões do modelo de referência está relacionada com uma das etapas do ciclo proposto por Clow [8], ilustrado na Figura 2. Este ciclo possui quatro etapas: (1) Alunos, (2) Dados, (3) Análise e (4) Intervenções.

A primeira etapa deste ciclo são os *alunos* que irão gerar os *dados* coletados na segunda etapa. Com estes dados, são realizadas diferentes análises que fornecerão conhecimento sobre o processo de aprendizagem. No entanto, o ciclo não se encerra enquanto os resultados destas análises não forem utilizados para direcionar *intervenções*, visto que são estas que tornam as ferramentas LA eficientes no aprimoramento do processo ensino-aprendizagem [8].

Brown [9] afirma que existem duas categorias de intervenções: as automáticas e as semiautomáticas. As intervenções automáticas são aquelas em que o sistema realiza ações pré-determinadas automaticamente, com pouca ou nenhuma participação do professor. Estas intervenções podem ser simples indicadores nas cores verde (bom desempenho), amarelo (desempenho regular) e vermelho (mau desempenho) alertando o aluno; ou ainda agentes inteligentes que oferecem ajuda e guiam o aluno diretamente nos assuntos em que encontram-se as deficiências de aprendizado. Por sua vez as intervenções semiautomáticas são alertas e outros indicadores relevantes dos padrões de aprendizagem enviados para os envolvidos no processo de ensino-aprendizagem (i.e., educadores e alunos). Estes padrões podem ser considerados como sintomas e os envolvidos podem decidir o que será feito para corrigir esta falha.

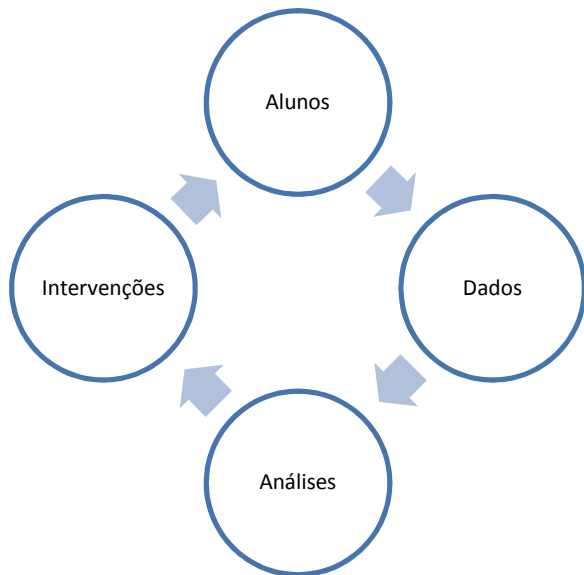


Figura 2. Ciclo de Learning Analytics [8]

Neste artigo adota-se uma versão adaptada da classificação de Brown (2012) — com o intuito de obter uma visão mais detalhada sobre como os educadores precisam se envolver para realizar as intervenções —, com as seguintes categorias: automáticas, semiautomáticas e manuais. A categoria *automática* refere-se às intervenções que não necessitam de participação do professor; a categoria *semiautomática* refere-se às intervenções com pouca participação do professor, ou seja, o professor apenas solicita que o sistema realize alguma tarefa; e a categoria *manual* refere-se às intervenções em que os envolvidos no processo de aprendizagem precisam tomar decisões individualmente ou em grupo e agir para alterar o processo.

Apesar de este trabalho possuir uma proposta similar ao trabalho de Atif *et al.* (2013) — que classifica quatro ferramentas *Learning Analytics* de acordo com as quatro dimensões definidas no modelo de Chatti *et al.* (2012), este trabalho abrange e classifica uma quantidade maior de artigos; analisa, além das dimensões do modelo de referências, as intervenções adotadas para melhorar o processo de ensino-aprendizagem; e faz a análise da relação entre dados, técnicas, público-alvo e intervenções *versus* objetivos.

3. METODOLOGIA

Esta seção detalha a metodologia da busca pelos trabalhos mapeados neste artigo. Na Seção 3.1 são descritas as questões de pesquisa, na Seção 3.2 é detalhado o processo de condução da busca pelos artigos e na Seção 3.3 é descrito como os artigos foram selecionados dentre os resultados da busca e como estes foram filtrados.

3.1. Questões de Pesquisa

Com base nas dimensões do modelo de referência de Chatti *et al.* (2012), no ciclo de Clow (2012) e na classificação de intervenções adaptada de Brown (2012), o mapeamento sistemático proposto neste artigo pretende responder às seguintes questões de pesquisa:

- Quais são os tipos de dados? Qual a relação destes com o objetivo da análise?
- Quais os métodos utilizados? Qual a relação entre os métodos e os objetivos?
- Quais os usuários para os quais os resultados foram destinados? Qual a relação do público alvo com o objetivo da análise?
- Quais as intervenções tomadas? Elas são automáticas, semiautomáticas, ou manuais? Qual a relação destas com o objetivo da análise?

3.2. Condução da Busca

A busca foi realizada nos Mecanismos de Busca Acadêmica (MBAs) IEEE Xplore, ACM, Scopus e Science Direct. Os MBAs Scopus e Science Direct foram escolhidos, pois estão entre os 10 mais bem avaliados em uma análise quantitativa de 40 MBAs [10]. Já os MBAs IEEE Xplore e ACM são escolhas justificáveis devido a sua relevância na área de Ciência da Computação. Para realizar a busca, foi utilizada a expressão booleana ilustrada na Figura 3, onde o asterisco (*) é um caractere coringa, ou seja, um caractere que pode ser substituído por outros N caracteres (onde $N \geq 0$).

("e-learning analytics" OR "elearning analytics" OR "learning analytics") AND (predict* OR intervention OR tutoring OR mentoring OR assessment OR feedback OR adaptation OR personalization OR recommendation OR "data mining" OR "social network analysis" OR "social analysis" OR "discourse analytics" OR "discourse analysis" OR "lexical analysis" OR dashboard* OR visuali*)

Figura 3. Expressão booleana utilizada na busca dos artigos

3.3. Critérios de Inclusão e Exclusão

Para facilitar a seleção de artigos, foram definidos alguns critérios de inclusão e exclusão. Os seguintes critérios de inclusão foram adotados:

- Artigos disponíveis na íntegra para *download*;
- Artigos de conferências ou periódicos;
- Artigos completos (quatro ou mais páginas);

Após realizar a busca nos mecanismos de busca definidos na Seção 3.2, foram pré-selecionados todos os artigos que satisfizessem os critérios de inclusão citados acima. Dos artigos pré-selecionados, foram excluídos todos aqueles que satisfizessem um ou mais dos critérios de exclusão a seguir:

- Não conter *learning analytics*, *elearning analytics* ou *e-learning analytics* no título, resumo, termos gerais ou palavras-chave;
- Artigos duplicados;
- Artigos que não mostram como realizar ou aplicar uma análise em uma ferramenta (artigos que não são primários), ou seja, artigos teóricos e de mapeamentos sistemáticos são desconsiderados.

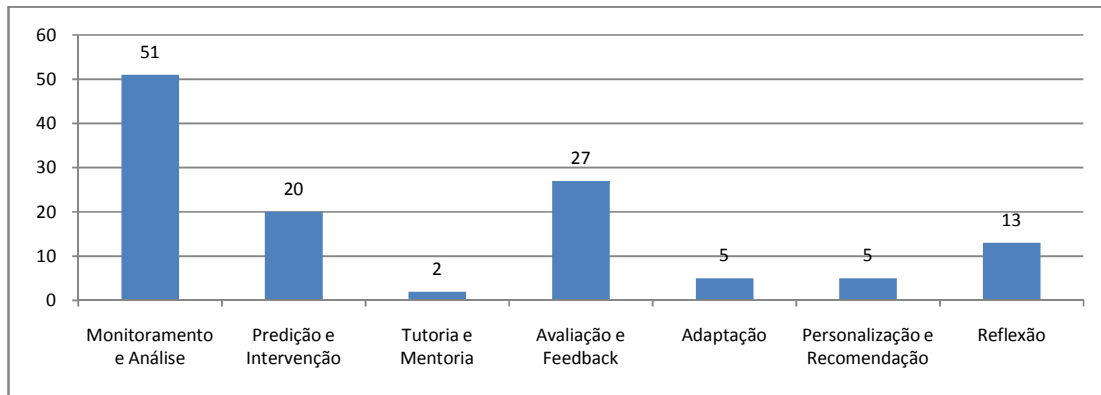


Figura 4. Distribuição dos trabalhos por objetivo

Após filtrar os artigos de interesse para este mapeamento, é importante ressaltar que foram considerados os trabalhos e não os artigos, assim caso um trabalho tenha sido descrito em mais de um artigo, os artigos foram considerados apenas como um. Desta forma, caso um trabalho ou partes dele tenham sido descritos em mais de um artigo, é possível obter informações mais detalhadas a respeito deste.

4. Resultados e Análises

Após realizar a busca conforme definido na Seção 3.2, foram pré-selecionados de acordo com os critérios de inclusão 642 artigos. Em seguida, foram aplicados os critérios de exclusão, eliminando 515 deles. Assim, foram selecionados 127 artigos, resultando em 116 trabalhos analisados (o mais antigo é de 2011 e o mais novo é de 2014). A Figura 4 mostra a distribuição destes trabalhos conforme as categorias da dimensão *Por quê?*, considerando que alguns destes possuem mais de um objetivo. Grande parte dos trabalhos está concentrada na categoria *Monitoramento e Análise*, ou seja, trabalhos em que são coletados os dados dos alunos para realizar análises. Os trabalhos das categorias *Avaliação e Feedback* e *Predição e Intervenção* também são foco de grande interesse na área.

Para alcançar os objetivos, os trabalhos utilizaram diferentes tipos de dados: anotações, áudio; dados demográficos (idade, gênero, classe social, etc.); dados contidos no ambiente de aprendizagem (metadados dos objetos de aprendizagem, publicações nos fóruns de discussão, etc.); dados coletados em entrevistas; dados do histórico escolar (notas, disciplinas cursadas, créditos obtidos, total de reprovações, etc.); dados de navegação dos alunos (páginas acessadas, tempo *online*, frequência de acesso, etc.) coletados nos ambientes de aprendizagem utilizados (Moodle, Blackboard, máquinas virtuais configuradas para monitorar as ações do aluno — dentro e fora da plataforma do curso —, plataformas para desenvolvimento de código, jogos sérios, etc.); dados coletados de mídias sociais (informações do perfil, comentários realizados, etc.) como YouTube, Fóruns de Discussão, etc.; dados coletados através de questionários; e dados obtidos através de vídeos.

A Figura 5 apresenta a relação entre os tipos de dados (representados nas linhas) e objetivos (representados nas colunas). A intensidade da cor da célula indica a quantidade de trabalhos encontrados (i.e. quanto maior a intensidade, mais trabalhos encontrados). Conforme a Figura 5, o tipo de dados predominantemente utilizado nas análises são os de navegação do aluno (50%). Os dados provenientes do AVA (12%) e os dados do

histórico escolar (13%) aparecem em uma quantidade significativa quando comparados aos demais tipos de dados (entre 1% e 7%). Dos demais tipos de dados, aqueles nos quais há certo interesse são os dados demográficos e os dados provenientes de mídias sociais.

Tipos dos dados		Objetivos							
		Monitoramento e Análise	Predição e Intervenção	Tutoria e Mentoria	Avaliação e Feedback	Adaptação	Personalização e Recomendação	Reflexão	Total
	Anotações	1	3		1				5
	Áudio	1	3		1				5
	Dados demográficos	3	9		1	1			14
	Dados do ambiente	8	1	1	9		1	2	22
	Entrevistas	1		1					2
	Histórico escolar	9	10		3	1	1		24
	Mídias sociais	3			3			5	11
	Dados Navegacionais	34	16	1	20	6	6	11	94
Questionários	3			2				5	
Vídeos	1	3		1				5	

Figura 5. Relação entre tipos de dados e objetivos

As análises da categoria *Monitoramento e Análise* — por serem mais exploratórias —, utilizam diversos tipos de dados, sendo o principal deles os navegacionais. Em análises de *Predição e*

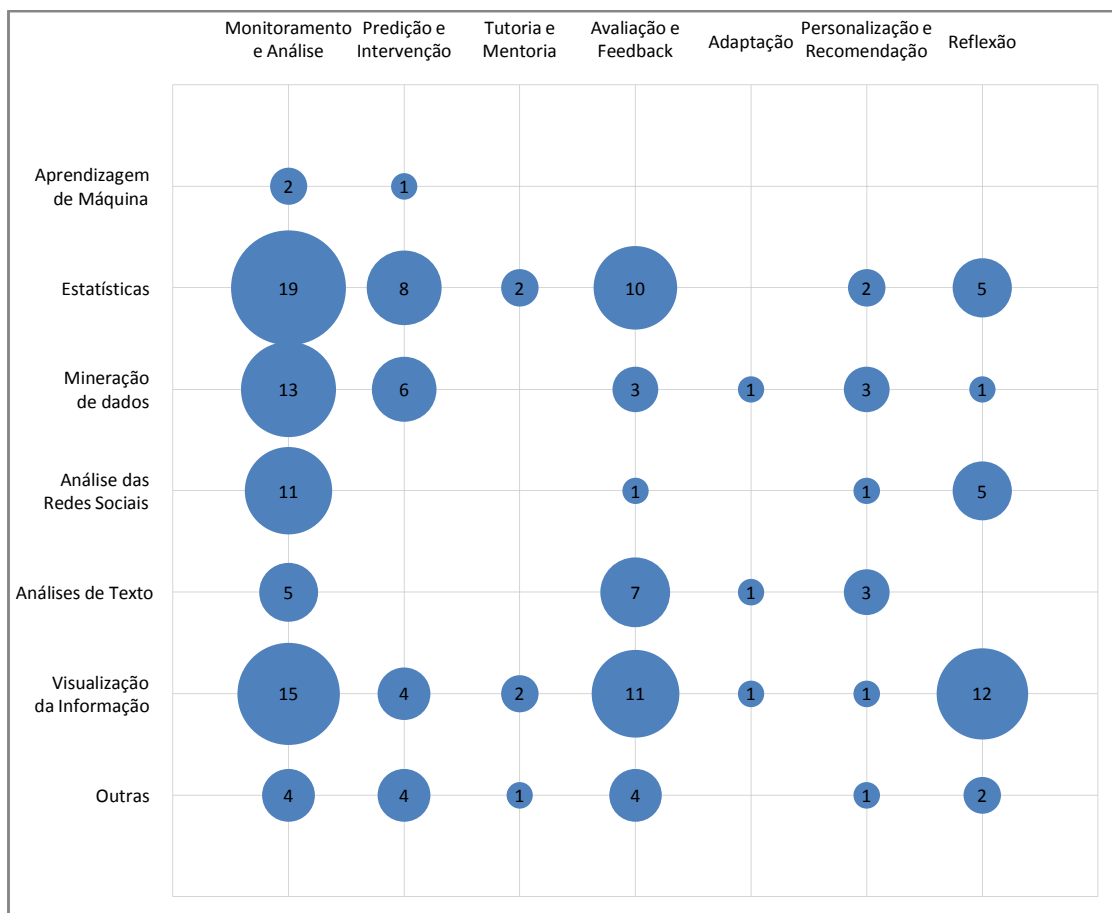


Figura 6. Relação entre técnicas e objetivos

Intervenção, os trabalhos focam principalmente nos dados demográficos, dados do histórico escolar e também nos dados de navegação do aluno. As análises de *Avaliação e Feedback* podem utilizar uma grande variedade de dados, prevalecendo os de navegação do aluno e os dados do ambiente.

Os dados são analisados através de técnicas de Aprendizagem de Máquina, Estatísticas, Mineração de Dados, Análise das Redes Sociais, Análises de Texto, Visualização da Informação, etc. No entanto, conforme pode ser observado na Figura 6, há a predominância de técnicas como Estatísticas e Visualização da Informação, aparecendo em grande quantidade e em grande parte das categorias. Técnicas de Mineração de Dados foram utilizadas para atingir todos os objetivos, exceto os da categoria *Tutoria e Mentoria* e aparecem em grande quantidade, juntamente com as técnicas de Análise de Redes Sociais, nos trabalhos da categoria Monitoramento e Análise.

Os resultados obtidos com estes trabalhos foram destinados a diferentes pessoas como administradores, alunos, conselheiros, desenvolvedores, professores e pesquisadores; podendo ser destinados a um grupo (por exemplo, uma ferramenta poderia possibilitar a reflexão tanto do aluno quanto do professor ao mesmo tempo). Conforme pode ser observado na Figura 7, grande parte dos trabalhos destinam-se aos pesquisadores, especialmente os da categoria *Monitoramento e Análise*. Como público final de uma ferramenta, estão em destaque os professores e os alunos.

	Objetivos						
	Monitoramento e Análise	Predição e Intervenção	Tutoria e Mentoria	Avaliação e Feedback	Adaptação	Personalização e Recomendação	Reflexão
Administradores	1						1
Alunos	4			13	4	5	8
Conselheiros		3				1	
Desenvolvedores		1	1	1			
Professores	11	3		16	1	1	6
Pesquisadores	34	14	1	6			
Total	55	18	2	26	5	7	15

Figura 7. Relação entre público alvo e objetivos

Após verificar os resultados das análises, conforme o ciclo de Clow (2012), é necessário realizar intervenções para melhorar o processo de ensino-aprendizagem. Nos trabalhos selecionados, as intervenções tomadas são adaptação no modo como o conteúdo é mostrado ao aluno, alterações pedagógicas, *feedback* ao aluno, orientações aos alunos, envio de notificações (SMS, e-mails, no AVA, etc.), recomendações de recursos e/ou colegas de trabalho; sendo que foram mapeados os trabalhos que identificavam claramente qual a intervenção que pode ser adotada e aqueles em que o trabalho claramente não possibilitava intervenções por explorarem indicadores, modelos, etc.

	Objetivos						
	Monitoramento e Análise	Predição e Intervenção	Tutoria e Mentoria	Avaliação e Feedback	Adaptação	Personalização e Recomendação	Reflexão
Intervenção							Total
Adaptação					6		6
Alteração pedagógica	2			3			2
Feedback				1			1
Orientações		3		1		1	6
Notificações		3		1			4
Recomendações						5	5
Sem intervenção	34	15	2	5			56

Figura 8. Relação entre intervenções e objetivos

As intervenções de adaptação, *feedback* e recomendação são automáticas; a alteração pedagógica e as orientações são manuais; e, por fim, o envio de notificações pode ser feito das três maneiras (automática, semiautomática e manual), dependendo do modo como os resultados são calculados, exibidos para o usuário final e do modo como o sistema foi desenvolvido/configurado.

Há poucos trabalhos que mencionaram a intervenção realizada e, no entanto, há uma grande quantidade de trabalhos não possibilitam alguma intervenção. Conforme a Figura 8, os trabalhos das categorias *Adaptação* e *Personalização e Recomendação* estão diretamente relacionados às intervenções *Adaptação* e *Recomendações*. Os trabalhos da categoria *Tutoria e Mentoria* estão relacionados à categoria *Sem intervenção*, sugerindo que este pode ser um aspecto a ser abordado na área de *Learning Analytics*.

A partir do mapeamento realizado nesta seção com artigos dos últimos 4 anos (2011-2014), nota-se um grande número de trabalhos da categoria *Monitoramento e Análise* (Figura 4), seguido pelo interesse em *Avaliação e Feedback* e *Predição e*

Intervenção. Há uma grande variedade de tipos de dados que são utilizados (predominando os de navegação) e também uma grande variedade de técnicas que podem ser aplicadas para analisar estes dados. Os resultados desta análise são direcionados para os alunos ou para os professores, possibilitando intervenções para melhorar o processo de ensino-aprendizagem. No entanto, os trabalhos quase não mencionam as intervenções, conforme observado na Figura 8.

5. MASSIVE OPEN ONLINE COURSES

Ao contrário dos AVAs (que servem para dar suporte à disciplinas presenciais ou a cursos *online*), os *Massive Open Online Courses* (MOOCs) surgem com o objetivo de massificar estes cursos [11]. Assim, os MOOCs são abertos ao grande número de usuários que desejam realizar um curso.

Os MOOCs criaram grandes oportunidades para pesquisadores e educadores interessados em LA. Devido ao grande número de alunos que participam deste tipo de curso, tem-se acesso a grandes conjuntos de dados das interações de alunos, possibilitando a compreensão das experiências, processos e resultados de aprendizagem [2]. Com o objetivo de melhor entender a área de LA para aplicar suas técnicas em um MOOC, a partir do mapeamento sistemático realizado, analisamos separadamente os trabalhos que analisam dados coletados durante um MOOC. Dos 116 trabalhos analisados, apenas 5 utilizam dados de MOOCs.

Utilizando dados provenientes da interação dos alunos em dois MOOCs distintos, Coffrin *et al.* [12] procuraram desenvolver técnicas de LA que podem ser utilizadas com dados provenientes de MOOCs e projetaram visualizações significativas para os usuários finais (instrutores, pesquisadores, estudantes e administradores) para descrever as interações dos alunos e os padrões de dedicação e sucesso. Também utilizando dados provenientes de MOOCs, Ferguson *et al.* [13] utilizam técnicas de linguística computacional (uma área interdisciplinar que utiliza estatísticas e modelagem da linguagem natural) para desenvolver e testar um *framework* que identifique o diálogo exploratório em discussões textuais para dar suporte a alunos e professores. Kizilcec *et al.* [14] utilizam dados de MOOCs para caracterizar a dedicação dos alunos em quatro grupos de acordo com suas características e comportamento. Através desses elementos eles identificam padrões de comprometimento e também os padrões da falta de comprometimento. Santos *et al.* [15] utilizam os dados para analisar como as atividades evoluem, se todas são relevantes e se o uso de um determinado idioma nos fóruns de discussão influenciam os resultados.

Também há trabalhos que já apresentam algum tipo de ferramenta. Ruipérez *et al.* [16] propõem uma ferramenta que permite a reflexão dos professores e alunos, além de recomendações de horários de estudo para os alunos (e.g., recomendar ao aluno o estudo pela manhã pois seu desempenho é maior) e recomendações para os professores de alunos que precisam de atenção especial (e.g. alunos que não acessaram mais o curso).

Em uma pesquisa bibliográfica tradicional conduzida sobre LA e MOOCs, outros três trabalhos se destacaram. O trabalho de Blanco *et al.* [17] propõe uma metodologia para criar um cMOOC (MOOC conectivista e baseado em redes sociais) adaptativo. Também foram propostas ferramentas que possibilitam a identificação de estudantes que podem vir a reprovar (predição) através de estatísticas [18] ou ainda com o uso de Cadeias de Ocultas de Markov [19].

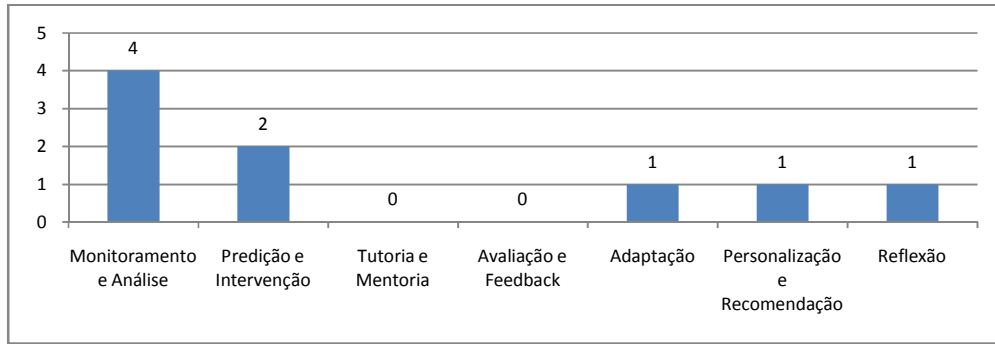


Figura 9. Distribuição dos trabalhos de Learning Analytics em Massive Open Online Courses por objetivo

Os oito trabalhos de LA em MOOCs foram distribuídos de acordo com seus objetivos na Figura 9 (assim como o gráfico da Figura 4, alguns trabalhos possuem mais de um objetivo). Conforme pode ser observado, não foram encontrados trabalhos que tratem da tutoria/mentoria e nem da Avaliação e *Feedback*. Em conjunto com a análise qualitativa, percebe-se que por mais que tenham sido encontrados trabalhos nas categorias Adaptação; Personalização e Recomendação; e Reflexão estes objetivos ainda possuem muitos ângulos a serem explorados.

6. DISCUSSÃO

Através do mapeamento sistemático realizado, pode-se verificar que a grande maioria dos trabalhos de LA está relacionado a categoria *Monitoramento e Análise*. No entanto, 56% dos trabalhos assim classificados continham análises exploratórias com propósitos diversos como, por exemplo, descobrir quais interações realizadas no ambiente de aprendizagem refletem positiva ou negativamente no aprendizado. De acordo com Duval e Verbert [20], este é um dos desafios da área devido ao pouco consenso sobre quais são os indicadores relevantes. Outros objetivos que a LA pode ajudar a atingir, mas ainda pouco explorados, são a tutoria, mentoria e recomendação.

Outro ponto a ser explorado é a diferença entre as ferramentas de avaliação; de *feedback*; de reflexão; e de monitoramento e análise. Salvo a ferramenta *Live Interest Meter* [21] que claramente possui características de *feedback* devido ao seu contexto de aplicação (i.e., o público avalia a apresentação e o apresentador recebe estes resultados em tempo real), as ferramentas inseridas nestas categorias coletam dados, calculam métricas/indicadores e os exibem para o que usuário final analise os resultados e obtenha suas conclusões. Alguns pontos a serem analisados são: Nestas ferramentas o que muda são as métricas/indicadores calculados? Estas ferramentas possibilitam que todos estes objetivos sejam alcançados de acordo com as necessidades dos usuários? Há alguma funcionalidade adicional, não relacionada diretamente a LA, que as diferencie?

Em relação a categoria *Personalização e Recomendação*, não há trabalhos que tratem especificamente da personalização. Por este motivo e pela definição do termo (i.e. configurar o ambiente de acordo com as preferências pessoais), acreditamos que este não é um objetivo de LA, mas um possível pré-requisito para a recomendação.

Em relação aos dados, os navegacionais são os que predominam. No entanto, a utilização desses dados nos trabalhos encontrados são basicamente em cursos tradicionais (para poucos alunos) e não em MOOCs (para muitos alunos). Dos trabalhos de LA que utilizam dados de MOOCs, alguns são exploratórios, outros apresentam propostas de ferramentas e apenas dois trabalhos apresentam ferramentas desenvolvidas. Neste contexto são

abordadas a predição, a reflexão, a adaptação e a recomendação. Devido a pouca exploração de LA em MOOCs e da grande quantidade de dados que podem ser coletados, confirmam-se as oportunidades de pesquisa.

A análise de dados como anotações, áudios e vídeos começa a ser explorada e pode permitir que a LA seja aplicada em salas de aula presenciais, visto que dados navegacionais, por exemplo, não podem ser coletados neste contexto e que apenas dados demográficos e de mídias sociais podem não ajudar a detectar o que ocorre neste ambiente. Além disso, dados coletados de mídias sociais também podem auxiliar na melhor compreensão do comportamento dos alunos e auxiliar no gerenciamento da aprendizagem colaborativa.

Técnicas de análise de texto começam a ser empregadas para investigar as atitudes dos alunos (e.g., se ele procura sanar uma dúvida ou ajudar outros alunos); técnicas de análise de redes sociais também são utilizadas para analisar a relação entre os alunos e seus colegas e recursos de aprendizagem. Ambas as técnicas podem fornecer novas informações sobre o contexto de aprendizagem.

A maioria dos trabalhos (47%) são de interesse de pesquisadores e não mostram ferramentas (onde o público-alvo são educadores e alunos). Considerando que quando a pesquisa mostra uma ferramenta o público-alvo não são os pesquisadores, nota-se na Figura 7 que é possível explorar ferramentas de tutoria e mentoria.

Há poucas intervenções sendo realizadas. Em parte porque 47% dos trabalhos analisados são destinados a pesquisadores e não a educadores e alunos; e em outra parte pois este elemento não foi abordado pelos autores. Assim, há um indício de que a intervenção é algo a ser melhor explorado no contexto das Learning Analytics.

7. CONSIDERAÇÕES FINAIS

As *Learning Analytics* são uma nova área de pesquisa na Informática na Educação que tem como objetivo melhorar o processo de ensino-aprendizagem através da análise de dados gerados pelos alunos. Estas análises podem ser feitas com diferentes tipos de dados (e.g., navegação do usuário, dados do AVA, dados demográficos, etc.) utilizando diversas técnicas (e.g., mineração de dados, estatísticas, visualização da informação, etc.) para atingir diferentes objetivos (e.g., predição, tutoria, adaptação, personalização, recomendação, etc.).

Os resultados destas análises podem ser mostrados para diferentes pessoas (e.g., alunos, professores, conselheiros, administradores, etc.) e podem acarretar em alguma alteração do processo de ensino-aprendizagem (e.g., modificação do sílabo e/ou materiais, orientações aos alunos, etc.). Estas alterações no processo ensino-aprendizagem são chamadas de intervenção e, conforme

mencionado por Clow (2012), são fundamentais para a efetividade de uma aplicação *Learning Analytics*. No entanto, nos resultados obtidos nesta pesquisa, poucos foram os trabalhos que mencionaram quais poderiam ser as intervenções aplicadas.

Esta é uma área nova e em expansão, com diversas pesquisas exploratórias sendo desenvolvidas com o objetivo de identificar quais são os dados relevantes (e.g., dados demográficos, dados navegacionais, dados acadêmicos, etc.) ou os modelos preditivos com a maior taxa de acertos, dentre outros.

Este trabalho apresentou um mapeamento sistemático sobre a área de *Learning Analytics* com intuito de entender seu conceito e como suas dimensões estão inter-relacionadas, bem como a área tem atuado e possíveis caminhamentos. O objetivo foi descobrir quais são os principais objetivos ao trabalhar na área, bem como analisar quais tipos de dados estão sendo utilizados, que técnicas são abordadas para cada objetivo, quais são as pessoas envolvidas bem como as intervenções realizadas. Também buscamos verificar o que está sendo feito na área em relação aos *Massive Open Online Courses*.

Acreditamos que a medida que a área evolui, outros objetivos serão propostos, bem como novas técnicas e intervenções serão elaboradas. Também acreditamos que há uma grande oportunidade para pesquisas na área em MOOCs, visto que há poucos trabalhos que abordem dados coletados nesta modalidade de Ensino a Distância.

8. AGRADECIMENTOS

Agradecemos à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC) pelo apoio durante a realização deste trabalho.

9. REFERÊNCIAS

- [1] LAK — 1st International Conference on Learning Analytics and Knowledge, Banff, Alberta, 27 fev-01 mar, 2011. Disponível em <<https://tekri.athabasca.ca/analytics/>>. Acesso em 06 mai 2014.
- [2] Siemens, G; Long, P. Penetrating the fog: Analytics in learning and education. *Educause Review*, v. 46, n. 5, p. 30-32, 2011.
- [3] Chatti, M. A.; Dyckhoff, A. L.; Schroeder, U.; Thüs, H. A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, v. 4, n. 5, p. 318-331, 2012.
- [4] Atif, A.; Richards, D.; Bilgin, A.; Marrone, M. Learning Analytics in Higher Education: A Summary of Tools and Approaches.
- [5] Goldstein, P.J. & Katz, R.N. (2005). *Academic Analytics: The Uses of Management Information and Technology in Higher Education*. ECAR Research Study, Vol. 8. Retrieved September 05, 2014, from <http://www.educause.edu/ECAR/AcademicAnalyticsTheUsesofMana/158588>.
- [6] International Educational Data Mining Society. Retrieved September 01, 2014, from <http://www.educationaldatamining.org/>.
- [7] Siemens, G., & d Baker, R. S. (2012, April). Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 252-254). ACM.
- [8] Clow, D. The learning analytics cycle: closing the loop effectively. In: *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. ACM, 2012. p. 134-138.
- [9] Brown, M. Learning analytics: moving from concept to practice. *Educause Learning Initiative*, 2012.
- [10] Buchinger, D., Cavalcanti, G. A. d. S., and Hounsell, M. d. S. (2014). Mecanismos de busca acadêmica: uma análise quantitativa. *Revista Brasileira de Computação Aplicada*, v. 6, n. 1, p. 108-120, 2014.
- [11] de Souza, M. V., Simon, R. M., Blanck, H. L., Ogliari, C. & de Matos, R. (2013). Mídias Sociais, AVAs e MOOCs: Reflexões Sobre Educação em Rede. *International Conference on Interactive Computer Aided Blended Learning* (pp. 183-190).
- [12] Coffrin, C., Corrin, L., de Barba, P., & Kennedy, G. (2014, March). Visualizing patterns of student engagement and performance in MOOCs. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge* (pp. 83-92). ACM.
- [13] Ferguson, R., Wei, Z., He, Y., & Buckingham Shum, S. (2013, April). An evaluation of learning analytics to identify exploratory dialogue in online discussions. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 85-93). ACM.
- [14] Kizilcec, R. F., Piech, C., & Schneider, E. (2013, April). Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proceedings of the third international conference on learning analytics and knowledge* (pp. 170-179). ACM.
- [15] Santos, J. L., Klerkx, J., Duval, E., Gago, D., & Rodríguez, L. (2014, March). Success, activity and drop-outs in MOOCs an exploratory study on the UNED COMA courses. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge* (pp. 98-102). ACM.
- [16] Ruipérez-Valiente, J. A., Muñoz-Merino, P. J., & Kloos, C. D. (2013, November). An architecture for extending the learning analytics support in the Khan Academy framework. In *Proceedings of the First International Conference on Technological Ecosystem for Enhancing Multiculturality* (pp. 277-284). ACM.
- [17] Blanco, Á. F., García-Peñalvo, F. J., & Sein-Echaluce, M. (2013, November). A methodology proposal for developing adaptive cMOOC. In *Proceedings of the First International Conference on Technological Ecosystem for Enhancing Multiculturality* (pp. 553-558). ACM.
- [18] Tabaa, Y., & Medouri, A. (2013). LASyM: A Learning Analytics System for MOOCs. *International Journal of Advanced Computer Science and Applications*, 4(5), 113-119.
- [19] Balakrishnan, G., & Coetzee, D. (2013). Predicting student retention in massive open online courses using hidden markov models.
- [20] Duval, E; Verbert, K. Learning analytics. *E-Learning and Education*, v. 1, n. 8, 2012. Disponível em <<https://eleed.campussource.de/archive/8/3336>>. Acesso em 09 jun 2014.
- [21] Rivera-Pelayo, V., Munk, J., Zacharias, V., & Braun, S. (2013). Live interest meter: learning from quantified feedback in mass lectures. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 23-27). ACM.