

EDMXP: uma ferramenta para mineração de dados em educação híbrida desenvolvida segundo a abordagem CCS¹

Rodrigo Yoshio Tamae
Universidade Federal do Vale do São Francisco (Univasf)
rodrigo.tamae@univasf.edu.br

Klaus Schlünzen Junior
Universidade Estadual Paulista (Unesp)
klaus@fct.unesp.br

RESUMO

A ampla utilização dos AVAs (Ambientes Virtuais de Aprendizagem) na Educação, podem contribuir para a utilização de TDIC (Tecnologias Digitais da Informação e da Comunicação), das Metodologias Ativas de Aprendizagem e que favorecem a abordagem CCS (Construcionista, Contextualizada e Significativa), na qual o cursista utiliza a tecnologia como instrumento para produzir algo que parte da sua vivência e realidade e o professor atua como mediador para ajudar este cursista a formalizar conceitos. Nesse contexto, a Internet e os dispositivos móveis têm contribuído para a proliferação de grande quantidade de dados em formato digital, mas ainda são pouco utilizados para gerar a descoberta de conhecimento. É onde se destaca a área de MDE (Mineração de Dados Educacionais), que consiste no desenvolvimento de métodos e técnicas orientados a explorar tais dados para melhor compreender o comportamento dos cursistas e em quais condições eles aprendem. O argumento norteador da pesquisa foi descobrir como as técnicas de MDE poderiam ser usadas para identificar indícios da abordagem CCS nos cursos da modalidade híbrida. O delineamento metodológico baseia-se na abordagem *Ex Post Facto*, pois a investigação foi realizada após a conclusão dos fatos. Para responder as questões norteadoras, o curso de pós-graduação em Educação Especial na Perspectiva Inclusiva do programa Redefor-Unesp foi analisado a partir das categorias CCS, sistematizados em um software experimental denominado EDMXP (*Educational Data Mining eXPeriment*). Os resultados foram compilados em uma linguagem que possibilita aos profissionais de Educação compreenderem melhor os dados para que, ao final, fosse possível constatar que a MDE pode ser um fator transformador em Educação, pois permite a tomada de decisão com base em dados e em fatos e não apenas de forma intuitiva ou por meio de experiências vivenciadas. Representa, portanto, uma nova forma de fazer e pensar a Educação.

Palavras-chave: Mineração de Dados Educacionais. Abordagem CCS. Educação híbrida.

¹ Este artigo é a versão expandida de uma comunicação oral apresentada no VII Simpósio de Educação Inclusiva e Adaptações/ V Simpósio Internacional de Educação a Distância. São Paulo, UNESP, 2019.

EDMXP: a data mining tool in hybrid education developed according to CCS approach

ABSTRACT

The wide use of VLE (Virtual Learning Environments) in Education can contribute to the use of ICT (Information and Communication Technologies), of Active Learning Methodologies and which favor the CCS approach (Constructionist, Contextualised and, Significant), in which the student uses technology as an instrument to produce something that is part of his experience and reality, and the teacher acts as a mediator to help this student formalize concepts. In this context, the Internet and mobile devices have contributed to the proliferation of large amounts of data in digital format, but they are still little used to generate knowledge discovery. This is where the area of EDM (Educational Data Mining) stands out, which consists of the development of methods and techniques oriented to explore such data to better understand the behavior of the students and under what conditions they learn. The research's guiding argument was to find out how EDM techniques could be used to identify evidence of the CCS approach in hybrid modality courses. The research design is based on the Ex Post Facto approach, as the research was carried out after the facts were concluded. To answer the guiding questions, the Redefor-Unesp Postgraduate Course in Special Education from an Inclusive Perspective was analyzed from the CCS categories, systematized in an experimental software called EDMXP (Educational Data Mining eXPeriment). The results were compiled in a language that enables education professionals to better understand the data so that, in the end, it would be possible to see that the EDM can be a transforming factor in education, because it allows decision making based on data and facts, and not just intuitively or through lived experiences. It represents, therefore, a new way of doing and thinking about Education.

Keywords: Educational Data Mining. Constructionist, Contextualized and Significant Approach. Hybrid Education.

EDMXP: una herramienta para la minería de datos en la educación híbrida desarrollada de acuerdo con el enfoque CCS

RESUMEN

El amplio uso de EVA (Entornos Virtuales de Aprendizaje) en educación, que puede contribuir al uso de TDIC (Tecnologías Digitales de Información y Comunicación), metodologías de aprendizaje activo y que favorecen el enfoque CCS (Constructivista, Contextualizado y Significativo), en el que el alumno utiliza la tecnología como

instrumento para producir algo que parte de su experiencia y realidad, y el profesor actúa como mediador para ayudarlo a formalizar conceptos. En este contexto, Internet y los dispositivos móviles han contribuido a la proliferación de grandes cantidades de datos en formato digital, todavía son pocos utilizados para generar el descubrimiento del conocimiento. Aquí es donde se destaca el área de MDE (Minería de Datos Educativos), que consiste en desarrollar métodos y técnicas destinados a explotar dichos datos para comprender mejor el comportamiento de los participantes del curso y en qué condiciones aprenden. El argumento rector de la investigación fue descubrir cómo las técnicas de MDE podrían usarse para identificar evidencia del enfoque CCS en cursos híbridos. El diseño metodológico se basa en el enfoque *ex post facto*, ya que la investigación se llevó a cabo después de la conclusión de los hechos. Para responder a las preguntas orientadoras, el curso de posgrado en educación especial en la perspectiva inclusiva del programa Redefor-Unesp se analizó a partir de las categorías CCS, sistematizado en un software experimental llamado EDMXP (*Educational Data Mining eXPeriment*). Los resultados se compilaron en un lenguaje que permite a los profesionales de la educación comprender mejor los datos para que, al final, sea posible verificar que el MDE puede ser un factor transformador en la educación, ya que permite la toma de decisiones basada en datos y hechos, y no solo de forma intuitiva o a través de experiencias vividas. Por lo tanto, representa una nueva forma de hacer y pensar sobre Educación.

Palabras clave: Minería de datos educativos. Enfoque CCS. Educación híbrida.

1 INTRODUÇÃO

Com a acentuada adoção das TDIC (Tecnologias Digitais da Informação e Comunicação), dados em formato digital foram e são produzidos em todo mundo em escala caótica e repetitiva, mas muitos destes dados, estruturados ou não, muitas vezes redundantes, permanecem confinados dentro das instituições, de dispositivos pessoais (móveis ou não) e outros se encontram espalhados pela Web. Mais dados não significam obrigatoriamente mais conhecimento disponível para a sociedade, pois é preciso implementar um meio de encontrar e entender o que é relevante e como fazer o uso inteligente dessas informações. A Educação, apesar de influenciada pela tendência de modernização e globalização, não conseguiu acompanhar esse movimento como fez a indústria, por exemplo.

Neste contexto, destacam-se os cursos de pós-graduação em Educação Especial e Inclusiva ofertados pelo Programa Redefor-Unesp (SCHLÜNZEN, 2015), que teve início em fevereiro de 2014 a partir de um convênio (registrado sob nº 9365/0400/2012) firmado entre a Unesp e a SEE-SP (Secretaria da Educação do

Estado de São Paulo), para o qual foram elaborados cursos de pós-graduação em Educação Especial na Perspectiva da Educação Inclusiva, tendo por público-alvo professores e gestores da rede pública estadual. O objetivo foi formar profissionais capazes de construir uma cultura colaborativa, capazes de articular a equipe escolar e os professores do SAPE (Serviço de Apoio Pedagógico Especializado), na perspectiva da Educação Especial e Inclusiva (SCHLÜNZEN, 2015).

Ofertados na modalidade híbrida, em formato acessível, modelados para uso em um Ambiente Virtual de Aprendizagem (AVA), estimularam a utilização de TDIC e das Metodologias Ativas de Aprendizagem, projetados segundo a abordagem Construcionista, Contextualizada e Significativa (CCS). Representam um cenário propício para a área de Mineração de Dados Educacionais (do original em inglês *Educational Data Mining*, EDM) devido à quantidade de dados em formato digital potencialmente passíveis de serem explorados para a descoberta de conhecimento (TAMAE, 2018).

Durante o processo de investigação para delineamento metodológico, os referenciais indicaram notória carência de pesquisas qualitativas (ROMERO *et al.*, 2011) e de ferramentas com recursos para reduzir a complexidade das tarefas de Mineração de Dados (do original em inglês *Data Mining* - DM) com potencial para serem utilizadas em contextos educacionais (COSTA *et al.* 2012; ROMERO, 2013; ROMERO e VENTURA, 2013) que usam recursos digitais em larga escala.

Estes elementos motivaram a implementação do protótipo da ferramenta denominada EDMXP (*Educational Data Mining eXPeriment*). No desenvolvimento desta ferramenta, dentre as várias tarefas de DM, foram utilizadas as de agrupamento (representadas pelos algoritmos Simple KMeans e VSM – *Vector Space Model*) e de classificação (representada pelo algoritmo J48), de tal forma que os resultados pudessem ser compilados em indicadores numéricos e representações gráficas, em uma linguagem que se supõe mais compreensível aos profissionais de Educação, de maneira que possam conduzir às mediações e ao acompanhamento das atividades de ensino e de aprendizagem no AVA com maior eficiência e eficácia.

2 OBJETIVO

Esta pesquisa teve como objetivo geral compreender as relações e as construções de produtos em cursos ofertados na modalidade híbrida e desenvolvidos segundo a abordagem CCS por meio de técnicas de DM, o qual se desdobra nos seguintes objetivos específicos: definir os principais aspectos conceituais e o delineamento metodológico; implementar técnicas para EDM em um protótipo de software para a identificação das categorias de análise da abordagem CCS; selecionar, coletar, preparar, processar e analisar um conjunto de dados para submetê-las às

tarefas de EDM; por fim, evidenciar o resultado e as discussões frente às teorias que fundamentam a abordagem CCS e a EDM.

3 FUNDAMENTAÇÃO

O delineamento teórico foi constituído a partir de três eixos: a DM, a EDM e a abordagem CCS. As principais referências de DM foram as obras de Witten, Frank e Hall (2011); Berry e Linoff (2004); e Han, Kamber e Pei (2012). No tocante à EDM, as principais referências foram os trabalhos de Romero (2013); Romero e Ventura (2010; 2013); Romero, Ventura, Penchenizkiy e Baker (2011); e Becker, Cummins, Davis, Freeman, Giesinger e Ananthanarayanan (2017). Para a abordagem CCS, as principais referências foram as teses de doutorado de Schlünzen (2000) e de Santos (2015), bem como a tese de Livre-docência de Schlünzen (2015).

De acordo com Berry; Linoff (1997, p.7), DM é a exploração e a análise de uma grande quantidade de dados, por meio automático ou semiautomático, a fim de descobrir padrões e regras significativos. Silva (2004) explica que as etapas de DM utilizam técnicas e algoritmos de diferentes áreas do conhecimento, principalmente IA (Inteligência Artificial), Banco de Dados e Estatística, conforme sugere a Figura 1.

Figura 1 - Áreas que formam a mineração de dados.



Fonte: Adaptado de Silva (2004).

Nas três áreas que contribuem para a formação da DM, há definições diferentes. Para a área de banco de dados, DM é a “extração automatizada ou de maneira conveniente de padrões que representam o conhecimento implicitamente armazenado ou recuperado em grandes bancos de dados, data warehouses, na Web, outros repositórios de informações maciças ou fluxos de dados” (HAN, KAMBER; PEI, 2012, p. XXIII). Na perspectiva da área de Inteligência Artificial, trata-se da “extração

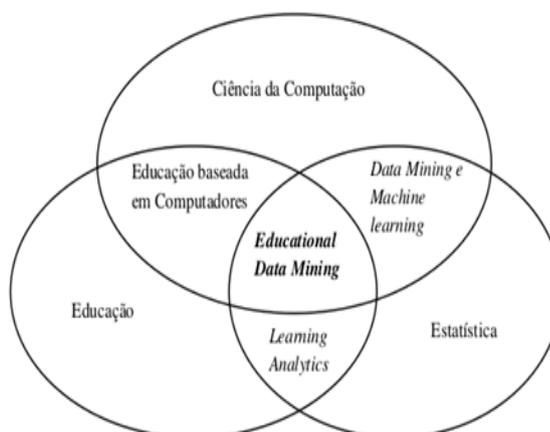
de informação implícita, previamente desconhecida e potencialmente útil a partir de dados” (WITTEN, FRANK e HALL, 2011, p. XXI). Para a estatística, DM é “um conjunto de funções matemáticas que descrevem o comportamento dos objetos em uma classe-alvo em termos de variáveis aleatórias e suas distribuições de probabilidade associadas a partir da coleta, análise, interpretação e apresentação de dados” (HAN, KAMBER e PEI, 2012).

Becker *et al.* (2017)² avaliam que há esforços perceptíveis da utilização de DM em sistemas de Educação on-line, na aprendizagem móvel e no AVA, de tal forma que, reunidos, possibilitem estruturar ambientes de análise e visualização para retratar dados de aprendizagem de forma multidimensional e móvel, pois cursos nas modalidades a distância e híbridos têm a capacidade de revelar como as ações dos estudantes contribuem para seu progresso e ganhos em termos de aprendizagem.

Como relatam Romero e Ventura (2013), a maioria das técnicas tradicionais de DM (classificação, agrupamento, mineração de relacionamentos, entre outras) foram testadas na Educação. Para Costa *et al.* (2012), a área de EDM procura desenvolver ou adaptar métodos e algoritmos de DM já existentes, de modo a ajudar a compreender melhor os dados em contextos educacionais, produzidos principalmente por estudantes e professores, nos ambientes nos quais eles interagem, normalmente, digitais.

Desta forma, é possível, por exemplo, avaliar e (melhor) entender o processo de aprendizagem do cursista e analisar sua interação com os recursos disponíveis (Costa *et al.*, 2012). A EDM tem suas raízes na combinação de três áreas principais, como pode ser observado na Figura 2. Note que, ao comparar com a DM (Figura 1), nota-se que a área da Ciência da Computação é representada pelas suas subáreas: Banco de Dados e Inteligência Artificial.

Figura 2 - Áreas que compõem a mineração de dados educacionais.



Fonte: Romero; Ventura (2013).

² The NMC Horizon Report: 2017 Higher Education Edition

Já na abordagem CCS, conforme explica Santos (2015), o cursista usa a tecnologia como instrumento ou recurso para produzir um produto do seu interesse (construcionista); as suas produções partem do seu próprio contexto, ou seja, a partir da sua vivência e realidade (contextualizada); ao construir esse produto, o cursista se depara com os conceitos das disciplinas curriculares e o professor atua como mediador na formalização e sistematização desses conceitos, para que o cursista atribua um significado aos mesmos (significativo).

No caso dos cursos do programa Redefor-Unesp, os cursistas foram estimulados a aprender a usar a tecnologia por meio do AVA; portanto, nesse contexto, estavam implícitos recursos como processadores de textos, gráficos, bancos de dados, objetos educacionais, sites, blogs, ferramentas de comunicação, ferramentas de colaboração, repositórios para compartilhar conteúdo digital e acessar informações e pessoas (SANTOS, 2015). Por isso, os recursos disponíveis no AVA possibilitam a sistematização das reflexões dos cursistas, a análise das suas reações, o compartilhamento de experiências, o acesso a materiais didáticos diversos e a reflexão sobre eles ao possibilitar o uso da tecnologia em seu contexto.

A partir desses aspectos fundamentais, nota-se a dificuldade de execução desta investigação, pois ressalta o desafio e expõe os pontos relevantes sobre as formas ou técnicas a serem utilizadas na verificação de como tais práticas podem realmente ser percebidas e devidamente identificadas nas ações de cursistas e professores no AVA.

4 METODOLOGIA E CONTEXTO DA INVESTIGAÇÃO

O delineamento metodológico *Ex Post Facto* (GIL, 2008) ou Pesquisa não experimental serviu de base para o estudo, pois foi realizado após a conclusão dos fatos. Para responder o problema de pesquisa, a turma “T1/14/EI” do curso de EEPI (Educação Especial na Perspectiva Inclusiva) do Programa Redefor-Unesp foi analisado a partir das categorias CCS definidas com base nas indicações de Schlünzen (2000; 2015), Santos (2015) e Valente (2005). O modelo conceitual de DM proposto por Fayyad, Piatetsky-Shapiro e Smyth (1996) serviu de base para o protótipo de software denominado EDMXP, implementado com tecnologias da plataforma Java para o suporte às atividades de seleção, pré-processamento, mineração e análise de dados.

Quanto à forma de abordagem científica, trata-se de uma pesquisa quali-quantitativa, pois teve a intenção de medir a relação entre as variáveis (quantitativa) e determinar opiniões ou projeções futuras nas respostas obtidas por meio de técnicas de EDM, bem como agregar valor às futuras conclusões baseado na premissa de que os problemas podem ser resolvidos e as práticas podem ser melhoradas por meio de descrição e análise de observações objetivas e diretas (qualitativa). Por exemplo,

levantar as opiniões, atitudes, crenças dos cursistas e a descoberta da existência de associações entre variáveis e a natureza dessas associações que possam apontar indícios da abordagem CCS.

5 RESULTADOS E DISCUSSÃO

As 5 (cinco) categorias de análise definidas, tendo por base a abordagem CCS, foram: Contexto do cursista, Espiral de aprendizagem e Ciclo de ações, Aprendizagem em rede, Papel do professor e Conceituação.

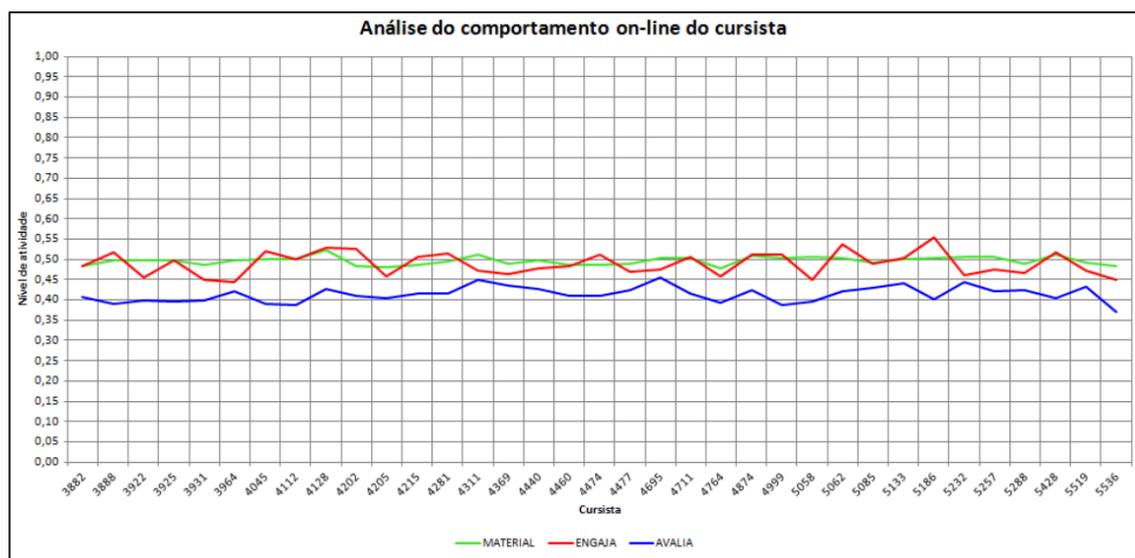
Os dados referentes à categoria de análise Contexto do cursista encontravam-se dispersos em tabelas do SGBD (Sistema Gerenciador de Banco de Dados) e o EDMXP possibilitou que os dados fossem sumarizados e consolidados para reduzir a dificuldade de análise de conteúdos e apontar indícios do contexto do cursista em sua produção.

Para isso, a investigação envolveu a relação entre duas atividades que ocorreram em fases distintas do curso: uma no início e outra no final. Na primeira atividade, os cursistas responderam a algumas questões para estreitar as relações entre eles (cursistas) e o TO (Tutor on-line) e forneceram dados a respeito de suas atividades pessoais e profissionais. Os cursistas se conheceram melhor ao interagir no fórum e iniciaram o processo de aprendizagem colaborativa em rede. Na atividade final, eles apresentaram um resumo de seus projetos, compartilharam suas experiências e sua trajetória metodológica. O elemento-chave para a consolidação destes dados foi a criação de uma funcionalidade no EDMXP para criação de categorias (*Tagset*) e rótulos (*tag*) específicos para anotar conjuntos de dados.

Nessa etapa, foi gerada a nuvem de palavras para auxiliar o processo de análise de conteúdo textual e que encontra-se seccionada em duas partes, cujo exemplo está representado na Figura 3: (A) Contexto e (B) Produto. Na parte A (lado esquerdo da figura), destacam-se as palavras “professora”, “atividades”, “respeito”, “ensino” e “escola”. Na parte B (lado direito da figura), destacam-se as palavras “educação”, “projeto”, “influência” e “relação”. Esses elementos podem indicar que o cursista, em seu projeto final, poderá contemplar uma educação norteadada pela influência da relação de afetividade entre professor e estudante.

Esse recurso facilita outras tarefas de análise como, por exemplo, a troca de mensagens que ocorre nos fóruns (Espiral de aprendizagem e Ciclo de ações, Aprendizagem em rede e Papel do professor). O fórum é o recurso que proporciona a todos os cursistas agirem como “iguais”, o que potencializa suas habilidades e qualidades (devido à sua característica assíncrona). Com esse recurso, pode-se visualizar as interações entre os cursistas e entre os cursistas e o TO, e assim, associar tais interações com os demais indicadores de EDM a partir de suas produções no AVA.

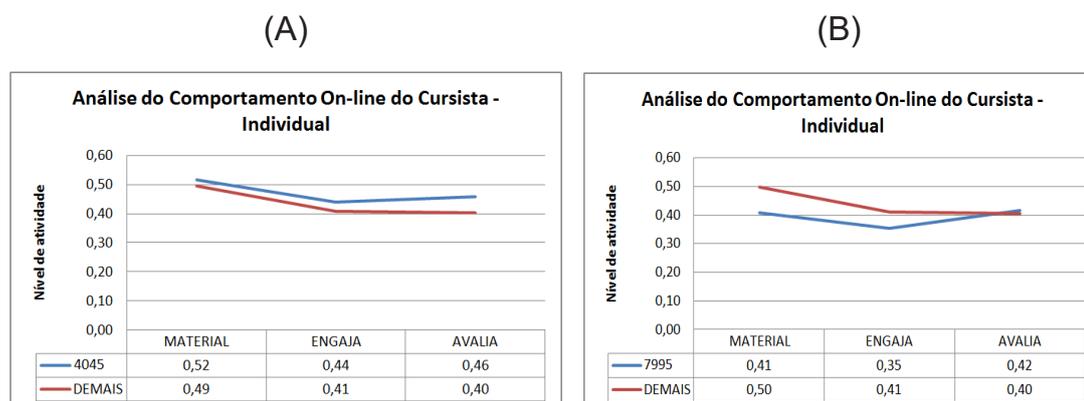
Gráfico 1 - Comportamento on-line do cursista.



Fonte: Autoria própria (2018).

A visualização do comportamento individual on-line é igualmente desejável para melhor analisar, acompanhar ou mesmo personalizar as atividades ao cursista. Um exemplo desta aplicação encontra-se representada no Gráfico 2, parte A (a esquerda), que revela o comportamento do cursista 4045 (em azul) em comparação ao comportamento médio dos demais (em vermelho), com base nas mesmas classes de atividades. No exemplo, pode-se ver que trata-se de um cursista que possui um desempenho acima da média. Em contrapartida, no exemplo representado no Gráfico 2, parte B (a direita), pode-se ver que o cursista 7995 (em azul) possui um desempenho abaixo da média quando comparado aos demais cursistas (em vermelho), mas apresenta melhor desempenho nas atividades de avaliação (um pouco acima dessa média). O que nos induz a questionar: o que isso significa? Havia algum erro nas atividades de avaliação que permitiram que este cursista pudesse burlar o sistema e ter bom desempenho nas atividades de avaliação? Ou não. O cursista apenas já conhecia previamente os elementos da atividade em questão e, por isso, o desempenho foi superior?

Gráfico 2 - Comparação do comportamento individual on-line.



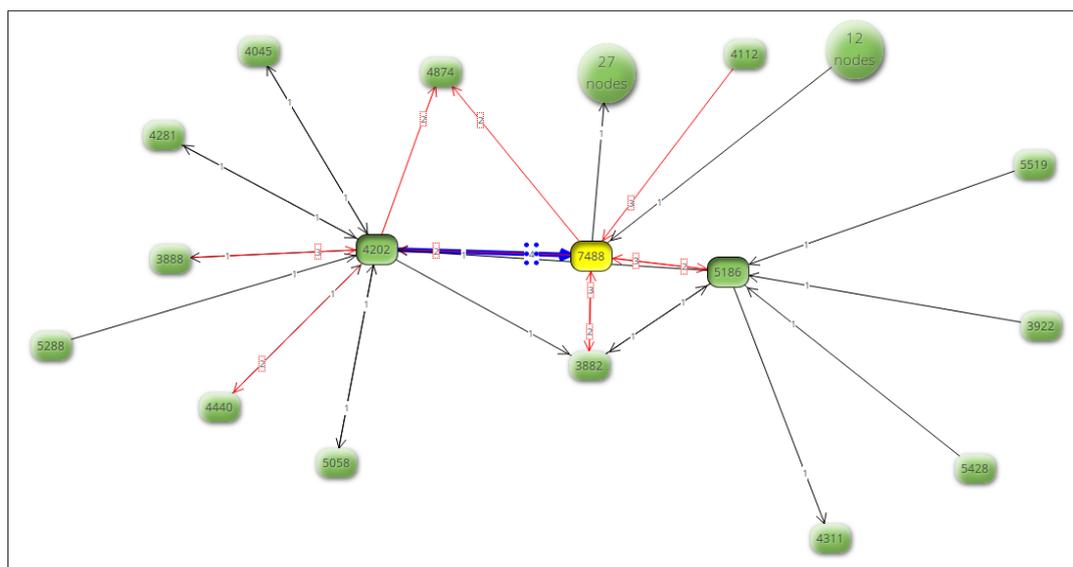
Fonte: Autoria própria (2018).

A Espiral de aprendizagem e o Ciclo de ações na abordagem CCS tem por objetivo desenvolver no cursista a habilidade de explorar, pesquisar, descrever, refletir e depurar seus pensamentos e formas de ação, e assim torná-lo capaz de produzir ou contribuir de forma significativa na execução de projetos que resultam nos produtos dos cursistas (SCHLÜNZEN, 2000; VALENTE, 2005; SCHLÜNZEN, 2015). Tal habilidade torna-se evidente em atividades de interação entre “cursistas e cursistas” e entre “cursistas e TO” (Aprendizagem em rede e Papel do professor).

Devido a isso, os grafos de interações representados na Figura 4 sugerem a necessidade de implementar meios de investigar o conteúdo dos fóruns com base na espiral de aprendizagem e do ciclo de ações para compreender sua contribuição para a construção do conhecimento que acontece nas interações dos cursistas (em verde) e tutores (em amarelo) por meio do AVA. A interação é representada por meio de linhas cujas cores e espessuras variam de acordo com a intensidade das trocas. Os grafos podem sinalizar aos professores autores se há necessidade de intervenções, bem como auxiliam na visualização de como o processo de mediação colaborativa ocorreu (SCHLÜNZEN, 2000; 2015).

Por isso, as postagens nos fóruns (interações) foram anotadas com um conjunto de quatro marcações disponíveis no EDMXP (1-Descreve, 2-Executa, 3-Reflete e 4-Depura), as quais permitem classificá-las de acordo com as etapas da Espiral de Aprendizagem e do Ciclo de ações (SCHLÜNZEN, 2000; VALENTE, 2005; SCHLÜNZEN, 2015) e oferece subsídios para a criação dos modelos de Mineração de Texto (do original em inglês *Text Mining* - TM).

Figura 4 - Grafo de interações no fórum.

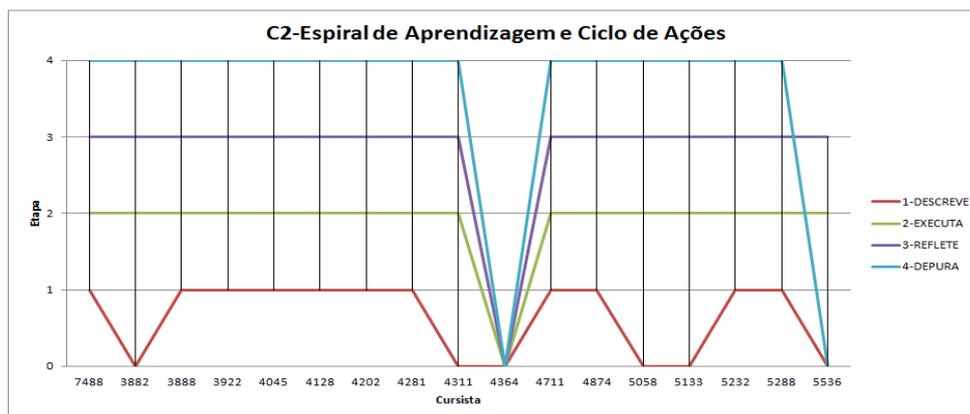


Fonte: Autoria própria (2018).

Se a DM trata a busca pelo reconhecimento de padrões em dados, do mesmo modo, a TM busca padrões em textos, pois é o processo de analisar textos e extrair informação que possam ser úteis em um contexto particular. No entanto, em termos de dados, o que os difere é que em TM, os dados não possuem uma estrutura e formato bem definidos, o que representa uma grande dificuldade de trabalhar com eles. Por se tratar de um dos principais veículos de informação, corresponde a uma tarefa que gera muita motivação, mas não se deve esquecer que, ao trabalhar com eles, o sucesso será apenas parcial (WITTEN, FRANK e HALL, 2011, p. 386).

Ao final, optou-se por construir um gráfico de linha com série de dados empilhadas, ilustrado no Gráfico 3, no qual as etapas “1-Descreve, 2-Executa, 3-Reflete e 4-Depura” (eixo Y) aparecem respectivamente nas cores vermelho, verde, roxo e azul. No eixo X estão representados os cursistas que compõem a amostragem. No exemplo, a maioria dos cursistas fazem uso de todas as etapas do ciclo de ações. No entanto, há os que não propuseram nenhuma solução (1-Descreve), mas participaram ativamente das demais etapas, o que pode indicar que puderam gerar ou colaborar para o melhor entendimento do que foi proposto ou mesmo sugerir novas formas de resolução do problema ou ainda apenas exploraram as ideias dos colegas e até as melhoraram.

Gráfico 3 - Etapas da Espiral de Aprendizagem.

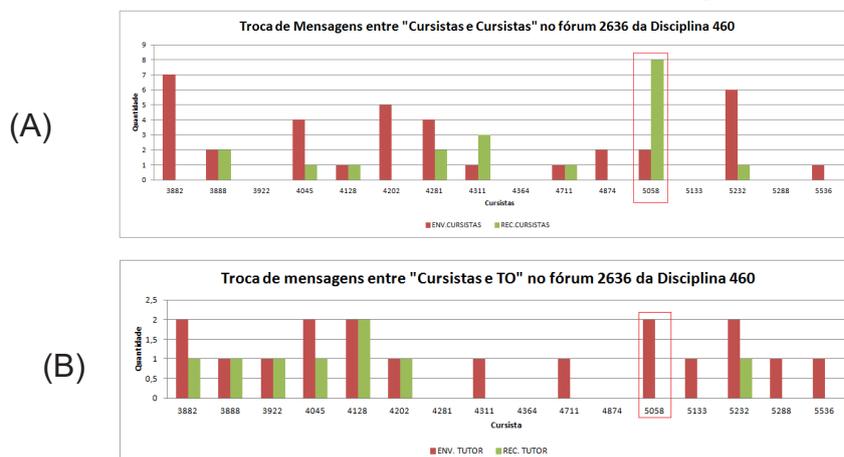


Fonte: Autoria própria (2018).

Essa é uma das razões pelas quais Schlünzen (2015, p.67) afirma que é por meio das redes de aprendizagem que os professores e cursistas aprendem a refletir sobre os fenômenos, dando significado e forma na construção do conhecimento, à medida que planejam, criam ações e trabalham colaborativamente.

O EDMXP também possibilita visualizar a quantidade de mensagens trocadas entre cursistas e entre cursistas e TO. No Gráfico 4, parte A (superior) estão representadas as trocas de mensagens entre cursistas. Na parte B (inferior) estão representadas as trocas de mensagens entre cursistas e o TO. O ponto forte deste resultado está na perspectiva qualitativa, pois conduz a vários questionamentos, como por exemplo, naqueles casos em que o cursista postou mais de uma mensagem, pode-se afirmar que ocorreu alto grau de engajamento ou trata-se de postagens que ocorrem em série e que, sozinhas, não apresentam sentido completo (indícios de fraude?) ou são postagens curtas (e sem sentido?). Há ainda aqueles que receberam mais postagens do que enviaram; seria isso indício de acomodação do cursista e os demais tentam instigá-lo a participar da rede? Ou sua postagem foi tão inovadora ou motivadora que fez com que os demais viessem a se manifestar?

Gráfico 4 - Exemplo de Troca de Mensagens.



Fonte: Autoria própria (2018).

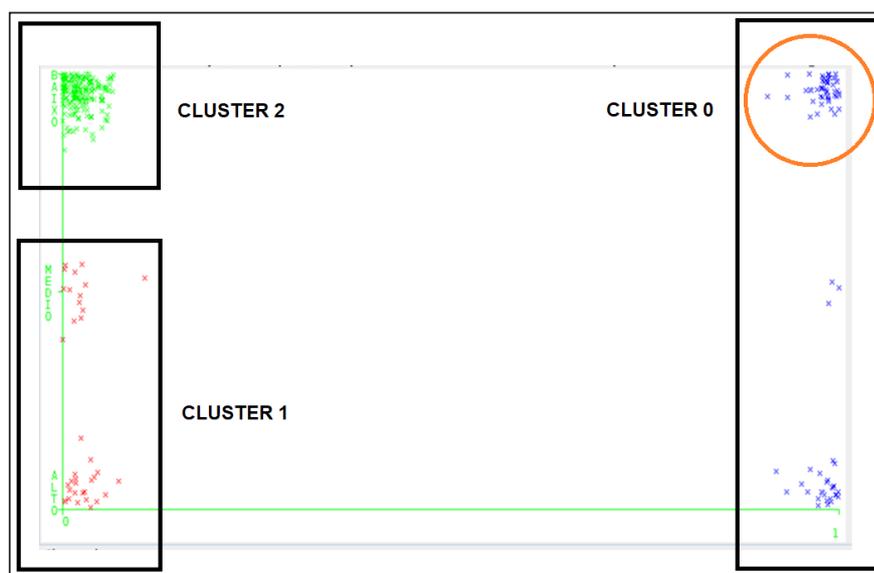
Na aprendizagem em rede, o papel do professor se destaca em uma abordagem CCS, pois a intensidade das trocas de mensagens entre o TO e os cursistas fornecem pistas quanto à sua capacidade em auxiliar os cursistas na espiral de aprendizagem e no ciclo de ações, na identificação dos momentos oportunos para auxiliar no processo de formalização de conceitos, bem como oferece a possibilidade de investigar a qualidade das postagens dos cursistas, ao ajudar a identificar, por exemplo, comportamentos distorcidos que podem ter impactos na produção do produto final.

Isso porque o que se espera em uma abordagem CCS é que o cursista seja capaz de se colocar em um estado de engajamento tal que possa construir um produto de seu interesse e não apenas cumprir processos para “obter nota e ser aprovado”.

Foi a motivação para classificar as postagens em alto (a partir da data de abertura do fórum), médio (no período médio entre as datas de abertura e encerramento) e baixo (próximo à data de encerramento do fórum) desempenho a partir do momento cronológico e do tamanho da postagem. Foram consideradas postagens curtas aquelas com menos de trinta palavras em sua composição e classificadas com ou sem sentido.

Na Gráfico 5, pode-se visualizar três agrupamentos nomeados automaticamente pelo algoritmo Simple KMeans como “cluster0”, “cluster1” e “cluster2”, representados, respectivamente, nas cores azul, vermelho e verde. O eixo X do gráfico demonstra a classificação das postagens como “Alto, Médio e Baixo” e o eixo Y apresenta os elementos “0” e “1”, no qual “0” representa a postagem normal (com mais de trinta palavras) e “1” representa a postagem curta (menor ou igual a trinta palavras). O que se pode verificar é que o “cluster0” (azul) é formado por postagens curtas geradas por cursistas que tiveram desempenhos classificados como “Alto”, “Médio” ou “Baixo”, produzidas durante todo o período em que o fórum esteve aberto. No entanto, como pode-se visualizar no destaque (elipse em laranja), há postagens suspeitas que podem (ou devem) ser investigadas. Já o “cluster1” (vermelho) foi formado por postagens de cursistas que tiveram desempenho “Alto” ou “Médio” e produziram postagens com mais de trinta palavras. Já o “cluster2” (verde) comporta postagens com mais de trinta palavras e desempenho “Baixo”, ou seja, postagens feitas no final do período de disponibilidade do fórum.

Gráfico 5 - Agrupamento das postagens curtas.



Fonte: Autoria própria (2018).

É neste cenário que ocorre a conceituação ou formalização de conceitos na abordagem CCS, principalmente nos momentos em que o professor interveio e naqueles em que o cursista engajado cumpriu as etapas do ciclo de ações em uma espiral contínua de aprendizagem. Assim, ao se deparar com uma situação de desconforto, frente ao conteúdo curricular proposto, a dúvida, o erro ou o sucesso parcial foram fundamentais para o processo de reflexão e depuração, pois foi o que colocou o cursista em uma condição de autonomia. Porém, como explica Valente (2005), nos momentos em que o cursista não pôde fazê-lo sozinho, ocorreu a intervenção dos pares ou do TO ou ainda dos professores autores em um contexto do EJV (Estar Junto Virtual).

6 CONSIDERAÇÕES FINAIS

Aplicar técnicas de EDM em cursos mediados pela abordagem CCS não é uma tarefa trivial, mas é possível afirmar que ao serem empregadas por meio do EDMXP, podem ser potencialmente utilizadas para gerar contribuições significativas no processo de desenvolvimento da abordagem CCS e possibilita aos pesquisadores uma nova forma de conduzir suas pesquisas, de maneira que concentrem foco mais em suas atividades de análise e reflexão, para apontar os aspectos definidos por Schlünzen (2015). Habilita o professor e o pesquisador a planejar e tomar decisão com base em dados, em fatos ocorridos, não apenas de forma intuitiva ou com base em experiências ou poucas informações contextuais.

Tecnicamente, deve-se considerar que é preciso empenhar grande esforço nos processos iniciais de DM, principalmente quanto às atividades de seleção e pré-processamento para a descoberta do conhecimento, pois viabilizam a construção de

modelos de DM, o que também justifica a implementação do protótipo EDMXP que os simplifica significativamente. Há ainda grande limitação tecnológica para a descoberta de conhecimento em estruturas textuais, o que é relevante, pois se constituem em um dos principais veículos de informação. Esta é a razão pela qual se faz necessário destacar que o sucesso nas tarefas que envolvem TM são (apenas) parciais.

Abstrair a complexidade das tarefas de DM para que possam ser utilizadas por profissionais sem o devido conhecimento na área de DM (ainda) representa um dos desafios amplamente reconhecidos na literatura. Por essa razão, o professor ou pesquisador que irá trabalhar efetivamente com a EDM está sendo concebido a partir de uma conjugação de fatores ou necessidades, um novo contexto fortemente influenciado pelo crescimento e proliferação das TDIC, e de dados em formato digital, mas não apenas isso. Será resultado de uma parceria verdadeira entre as diversas áreas da ciência.

REFERÊNCIAS

BECKER, S.A.; CUMMINS, M.; DAVIS, A.; FREEMAN, A.; GIESINGER, C.H., ANANTHANARAYANAN, V. **NMC Horizon Report: 2017 Higher Education Edition**. Austin, Texas: The New Media Consortium, 2017.

BERRY, M. J.A., LINOFF, G. **Data mining techniques: for marketing, sales and customer relationship management**. 2.ed. Willey, 2004.

COSTA, E.; BAKER, R. S.J.d.; AMORIM, L.; MAGALHÃES, J.; MARINHO, T. Mineração de Dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações. **Anais da 1..Jornada de Atualização em Informática na Educação / Seiji Isotani, Fernanda C. A. Campos**. Rio de Janeiro: Universidade Federal do Rio de Janeiro, 2012.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P.. **From data mining to knowledge discovery: An overview**. *In: Advances in Knowledge Discovery and Data Mining*, AAAI Press/The MIT Press, England, 1996, p.1-34.

GIL, A.C. **Métodos e técnicas de pesquisa social**. 6.ed. São Paulo:Atlas, 2008.

HAN, J.; KAMBER, M.; PEI, J. **Data Mining Concepts and Techniques**. 3.ed. Elsevier, 2012.

ROMERO, C. Minería de Datos en Educación y Análisis del Aprendizaje. Video. *In: Las Jornadas de Learning Week (LASI Spain)*, Salón de Actos de la Facultad de Ciencias del Trabajo de Granada, 04 de julio de 2013, 11:45 horas. Disponível em: https://www.youtube.com/watch?v=nAo5jQ_3H9k. Acesso em: 20 ago. 2016.

ROMERO, C.; VENTURA, S. Data mining in education. **WIRES Data Mining Knowledge Discover**, v.3: 12–27. 2013. doi:10.1002/widm.1075

ROMERO, C.; VENTURA, S.; PECHENIZKIY, M.; BAKER, R. S. de J. (eds.). **Handbook of Educational Data Mining**. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. CRC Press, 2011.

SANTOS, D.A.N. **A Abordagem CCS na Formação de Professores Para Uma Escola Inclusiva**. Tese (Doutorado) - Unesp - Presidente Prudente-SP. 2015.

SILVA, M. P. S. **Mineração de Dados: conceitos, aplicações e experimentação com weka**. ERI RJ/ES IV. Vitória – ES / Rio das Ostras / RJ. Novembro, 2004. Disponível em: <http://www.lbd.dcc.ufmg.br/colecoes/erirjes/2004/004.pdf>. Acessado em: 20 abr. 2016.

SCHLÜNZEN, E. T. M. **Abordagem Construcionista, Contextualizada e Significativa: formação, extensão e pesquisa em uma perspectiva inclusiva**. Tese (Livre-Docência) - Unesp - Presidente Prudente-SP. 2015.

SCHLÜNZEN, E. T. M. **Mudanças nas Práticas Pedagógicas do Professor: Criando Um Ambiente Construcionista Contextualizado e Significativo para Crianças com Necessidades Especiais Físicas**. Tese (Doutorado) - PUCSP - São Paulo. 2000.

TAMAE, R.Y. **Técnicas de mineração de dados em educação híbrida desenvolvida segundo a abordagem ccs**. Tese (Doutorado) – Unesp – Presidente Prudente-SP. 2018.

VALENTE, J. A. **A Espiral da Espiral de Aprendizagem: o processo de compreensão do papel das tecnologias de informação e comunicação na Educação**. Tese (Livre-Docência) - Universidade Estadual de Campinas. Instituto de Artes. Campinas-SP. 2005.

WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data Mining: practical machine learning tools and techniques**. Elsevier, 2011.

BIOGRAFIA DOS AUTORES

RODRIGO YOSHIO TAMAE – Doutor em Educação pela FCT/Unesp, Mestre em Ciência da Computação pelo UNIVEM e Graduado em Processamento de Dados pela Unimar. Professor do Curso de Ciência da Computação da Univasf.

KLAUS SCHLÜNZEN JUNIOR – Livre-docente em Informática e Educação pela Unesp, Pós-doutorado pela Universitat de Barcelona e Becário da Fundación Carolina, Doutor em Engenharia Elétrica pela Unicamp, Mestre em Ciência da Computação pela Unicamp e Licenciado em Matemática pela UFSC. Professor do Departamento de Estatística, do PPGE da FCT/Unesp, Coordenador geral do PROFEI e Coordenador do CPIDES.

Data de recebimento: 20/09/2020

Data de aprovação: 23/10/2020