

# Metodologias de Mineração de Dados aplicadas a Ambientes Educacionais Online

Geraldo Ramos Falci Júnior, Ivan L. M. Ricarte (Orientador)

Departamento de Engenharia de Computação e Automação Industrial (DCA)  
Faculdade de Engenharia Elétrica e de Computação (FEEC)  
Universidade Estadual de Campinas (Unicamp)  
Caixa Postal 6101, CEP 13083-970 – Campinas, SP, Brasil

{falci,  
ricarte}@dca.fee.unicamp.br

**Abstract** – Distance education brings about many difficulties for instructors to understand their classes needs and thus to improve their courses content. We analyze some methodologies to evaluate online courses, focusing on approaches to locate and retrieve information about content usage. We explore the possibilities of data mining for processing this information, searching for useful usage patterns that may help instructors to identify problematic spots on their contents. We evaluate the application of these methods on information gathered from online courses administered at Unicamp using the TIDIA-Ae software. We relate our achievements so far and our expectations for future developments of this work.

**Keywords** – data mining, distance education, online education, e-learning, TIDIA-Ae, Sakai.

para que tais conteúdos tenham acessibilidade facilitada.

## 1. Introdução

Em toda forma de ensino acadêmico é um desafio do educador entender as percepções e reações dos alunos ao curso ministrado. A partir desse entendimento, o professor consegue ampliar e aperfeiçoar o curso, revisando o material repassado aos alunos e a melhor forma de apresentá-lo. Na educação tradicional em classe, muito desse entendimento vem do contato direto com os alunos. Mas no ensino à distância e nas interações assíncronas, esse contato em geral se perde. É necessário obter tais informações de outra maneira.

No caso específico de cursos online, parte das informações necessárias ao professor podem estar ocultas em meio a uma enormidade de dados de acesso e registros de atividades dos alunos. Técnicas de mineração de dados podem ser aplicadas para localizar, filtrar e organizar informações a partir desses dados, permitindo assim prover, da forma mais simples e direta possível, um retorno ao professor ou autor de conteúdos sobre o uso que está sendo feito do material disponibilizado.

Milhares de cursos são lançados todo ano. A maioria destes consiste em redes de páginas estáticas, o que pode levar a dificuldades de orientação e compreensão para os estudantes, uma vez que a navegação nessas redes é irrestrita [5]. Uma possível consequência é a possibilidade de haver porções do conteúdo que dificilmente sejam visitadas pelos alunos, necessitando uma revisão na organização da rede

Nosso trabalho explora as possibilidades de descoberta desse tipo de informação por meio da aplicação de técnicas de mineração de dados a registros de utilização de cursos disponibilizados pela Web. Para fins de validação das propostas será utilizado o ambiente TIDIA-Ae, software que está sendo avaliado pelo Centro de Computação da Unicamp para substituir o ambiente TelEduc como plataforma de apoio online para os cursos de graduação e pós-graduação da Unicamp. Os resultados deste trabalho poderão ser usados não apenas por professores e autores, conforme já indicado, mas também por desenvolvedores do ambiente, que poderão incluir nas ferramentas desenvolvidas a coleta de informações que poderão melhorar a qualidade dos resultados obtidos no processo de mineração dos dados.

## 2. Aplicação de técnicas de mineração de dados para cursos na Web

Mineração de dados refere-se genericamente ao processo de extrair conhecimento a partir de grandes volumes de dados. Tal processo ocorre por meio da aplicação iterativa dos seguintes passos [3]:

1. Limpeza dos dados: remoção de ruídos e inconsistências nos dados;
2. Integração dos dados: combinação de múltiplas fontes de dados;
3. Seleção dos dados: recuperação dos dados que são relevantes para a análise;

4. Transformação dos dados: consolidação dos dados na forma apropriada para a mineração, por meio da aplicação de operações tais como sumarização e agregação;
5. Mineração dos dados: aplicação de métodos baseados em técnicas da área de inteligência computacional para a descoberta de padrões nos dados;
6. Avaliação de padrões: identificação de padrões de interesse por meio de alguma métrica apropriada à aplicação;
7. Apresentação do conhecimento: utilização de técnicas de visualização e representação de conhecimento para apresentar o conhecimento extraído ao usuário.

### 2.1. Mineração de Dados para a Web

Embora a principal área de aplicação das técnicas de mineração de dados ainda seja na área comercial, muitas outras áreas que também lidam com grandes volumes de dados vêm cada vez mais utilizando essas mesmas estratégias para outras aplicações. A Web, com seu enorme volume de recursos e usuários, é naturalmente uma fonte de dados para um grande número de aplicações.

Há duas abordagens que são usualmente utilizadas para obter os dados que alimentam o processo de mineração de dados em aplicações Web. Algumas das soluções desenvolvidas obtêm dados interferindo diretamente sobre os recursos, acrescentando pequenos programas que capturam e armazenam informações de uso, como realizado por Cardieri [1]. Outras buscam interferir minimamente com os conteúdos, geralmente aproveitando os registros de uso (*logs*) presentes nos servidores Web, como Zaiane e Luo [8].

Considerando os tipos de dados que estão presentes na Web, as quatro principais áreas de aplicação das técnicas de mineração de dados para a Web são [6]:

- **Mineração de conteúdo**, usada para descobrir informações diversas a partir do conteúdo das páginas Web, principalmente (mas não limitada a) textos e gráficos.
- **Mineração estrutural**, que analisa a organização do conteúdo. Pode ser intrapágina, quando utiliza a informação da estrutura interna do documento (por meio da análise de rótulos HTML ou XML) ou inter-páginas, quando foca nos

hyperlinks e na forma como se conectam diversas páginas.

- **Mineração de uso**, que observa a exploração de dados gerados pelos acessos dos usuários aos recursos, tais como origem da solicitação, referências solicitadas e momento dos acessos.
- **Mineração de perfil de usuário**, que explora informações demográficas sobre os usuários da Web por meio de dados de registros e informações de perfis disponibilizadas pelos usuários.

### 2.2. Mineração de Dados aplicada a Sistemas de Ensino Online

A mineração de dados, quando aplicada a sistemas de ensino, frequentemente é apoiada nos usos das mesmas técnicas às aplicações comerciais, fazendo um paralelo entre a análise do comportamento dos alunos enquanto navegam pelas páginas do curso com a navegação de um consumidor pelas páginas de produtos de uma empresa. De fato, alguns pesquisadores, como Zaiane [7], propõem experimentar diretamente as técnicas comprovadas em sites comerciais para cursos online. No entanto, reconhecidamente aplicações educacionais oferecem possibilidades mais amplas, fazendo com que a área de mineração de dados educacionais venha crescendo de forma significativa [4].

A busca por padrões em dados coletados de ambientes educacionais ocorre principalmente pela descoberta de *regras de associação* ou pela aplicação de técnicas de *classificação* ou de *agrupamento* [5]. O primeiro tipo visa discernir padrões de acesso às páginas dos cursos pelos alunos ou encontrar associações entre as páginas visitadas. O segundo objetiva agrupar alunos pelo comportamento de acesso que apresentam, buscando similaridades entre eles.

Em todos os casos, os resultados do processo de mineração de dados são aplicados e processados para chegar a conclusões diversas, num foco comum de melhorar aspectos diferentes do aprendizado em cada curso. Algumas áreas de aplicação da mineração de dados em aplicações educacionais incluem a avaliação de desempenho de alunos, adaptação e recomendação de conteúdos com base no comportamento de alunos, a detecção de comportamentos de aprendizagem atípicos por parte dos alunos, a avaliação dos conteúdos educacionais e a disponibilização de retorno (*feedback*) a alunos e instrutores [2].

Neste trabalho, a técnica de mineração de uso será utilizada para realizar a análise de conteúdos educacionais disponibilizados na Web, com o objetivo de oferecer retorno a instrutores e autores de conteúdos educacionais.

### 3. Desenvolvimento

A metodologia para a execução deste trabalho adotou como ponto de partida, além da revisão bibliográfica na área de mineração de dados educacionais, um estudo de caso. Para a realização desse estudo de caso, o primeiro passo foi selecionar um ambiente educacional utilizado para disponibilizar os documentos educacionais.

O ambiente selecionado foi o TIDIA-Ae, desenvolvido no contexto do projeto Tecnologias da Informação para Desenvolvimento da Internet Avançada, da FAPESP [9]. A versão atual desse ambiente foi desenvolvida sobre a plataforma SAKAI [10]. O ambiente TIDIA-Ae, na versão desenvolvida pelo grupo da Unicamp, será provavelmente adotado pela Unicamp como a principal plataforma para apoio aos cursos de graduação e de pós-graduação, já tendo sido adotado pela Universidade Virtual do Estado de São Paulo (Univesp). Além desses fatos, o nosso envolvimento no projeto TIDIA-Ae permite que recomendações resultantes deste trabalho possam ser incorporadas em versões futuras do ambiente.

Definido o ambiente, o próximo passo foi avaliar as possíveis fontes de dados para o processo de mineração. Os registros de acesso dos usuários, mantidos pelos servidores de aplicação, geralmente são fontes significativas de informações. No entanto, eles provêm apenas um traçado rudimentar das navegações e atividades do aprendiz no ambiente; uma etapa significativa de limpeza e transformação deve ser realizada para que a informação possa ser utilizada pelos algoritmos de mineração de dados [7]. Outras possíveis fontes são os registros específicos associados a ferramentas disponibilizadas no ambiente, mas neste caso foi preciso considerar que nem todas as ferramentas oferecem tal recurso.

Com a definição da fonte de dados, o estudo de caso servirá para avaliar o esforço necessário para realizar a extração e limpeza dos dados, bem como para avaliar qual o tipo de informação poderá ser obtida desses dados e a melhor forma de processá-los. Também serão avaliadas as estratégias para armazenamento desses dados processados, provavelmente utilizando um sistema gerenciador de banco de dados relacional para este fim.

O resultado desse estudo de caso permitirá que seja desenvolvida uma proposta para uma aplicação com o objetivo específico de prover o retorno para instrutores e autores de conteúdos educacionais sobre o uso dos materiais disponibilizados por meio do ambiente TIDIA-Ae. Para validar a proposta deste trabalho, bem como a aplicação desenvolvida, novos registros – desta feita, com o conhecimento prévio do tipo de informação que deve ser mantida – serão obtidos e utilizados.

### 4. Resultados

Para realizar o estudo de caso, obtivemos junto aos administradores das aplicações de EAD do Centro de Computação da Unicamp os registros dos acessos ao servidor de aplicação (tomcat) relativos aos cursos apoiados pelo TIDIA-Ae no segundo semestre de 2008. Uma primeira avaliação demonstrou que essa fonte de dados limita a qualidade da informação obtida, pois desconsidera a estrutura interna das ferramentas do ambiente.

Como a plataforma Sakai permite a manutenção de registros de eventos internos, essa é potencialmente outra fonte de dados a ser considerada. Tais registros, mantidos em uma tabela (SAKAI\_EVENT) da base de dados do sistema, são provisoriamente mantidos para fins de manutenção pelos desenvolvedores e pode, para fins do estudo de caso, oferecer subsídios adicionais, embora dificilmente essa fonte possa ser considerada como uma possibilidade para a aplicação final. Ao momento da finalização deste artigo, ainda aguardamos a liberação de acesso a essa fonte de dados para averiguar as informações presentes e dar prosseguimento ao estudo de caso.

### 5. Conclusões

Neste artigo apresentamos algumas noções sobre aplicações de técnicas de mineração de dados em ambientes de educação online com o intuito de fornecer um instrumento de auxílio ao educador para corrigir problemas e melhorar os cursos oferecidos à distância.

Discutimos o trabalho que realizamos até agora e as expectativas em relação aos próximos passos planejados.

Apesar de ter nosso desenvolvimento focado para o ambiente TIDIA-Ae, acreditamos que os resultados obtidos poderão facilmente ser adaptados para outros ambientes de aprendizado online. Além de prover uma ferramenta útil para

instrutores e autores de conteúdos educacionais, os resultados deste trabalho oferecerão retorno para indicar, aos desenvolvedores de ferramentas educacionais, qual o tipo de informação que deve ser mantida para oferecer um retorno de boa qualidade.

## Referências

- [1] Maria Angélica C. de Andrade Cardieri. *Mecanismo de monitoramento do uso de recursos Web para apoio à avaliação de ambientes*. Dissertação de mestrado, FEEC-Unicamp, 2004.
- [2] Félix Castro, Alfredo Vellido, Ângela Nebot e Francisco Mugica. *Applying Data Mining Techniques to e-Learning Problems*. *Studies in Computational Intelligence (SCI)*, 62, 183-221, 2007.
- [3] Jiawei Han e Micheline Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, 2001.
- [4] C. Romero e S. Ventura. *Educational Data Mining: A Survey from 1995 to 2005*. *Expert Systems and Applications*, 33, 135-146, 2007.
- [5] Cristóbal Romero, Sebastián Ventura e Paul De Bra. *Knowledge Discovery with Genetic Programming for Providing Feedback to Courseware Authors*. *User Modeling and User-Adapted Interaction*, 14: 425-464, Dezembro 2004.
- [6] Jaideep Srivastava, Robert Cooley, Mukund Deshpande e Pang-Ning Tan. *Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data*. *SIGKDD Explorations* 1(2), 12-23, Janeiro 2000.
- [7] Osmar R. Zaiane. *Building a Recommender Agent for e-Learning Systems*. *Proceedings of the IEEE International Conference on Computers in Education*, 55-59, 2002.
- [8] Osmar R. Zaiane e Jun Luo. *Towards Evaluating Learners' Behaviour in a Web-Based Distance Learning Environment*. *Proceedings of the IEEE International Conference on Advanced Learning Technologies*, 357-360, 2001.
- [9] Sobre o Projeto – Tidia-AE. <http://tidia-ae.incubadora.fapesp.br/portal/o-projeto>. (acessado em 11/02/2009).
- [10] Sakai Project: Home: About Sakai. <http://sakaiproject.org/portal/> (acessado em 14/02/2009).