

# Uso de Técnicas de *Data mining* no Monitoramento de Alunos *On-line*\*

Fabrcio Rangel Cunha  
Rafael Lucena Pessoa de Araujo  
Neide dos Santos  
neide@ime.uerj.br

## Resumo

*A Internet oferece amplas possibilidades para a criaçao de ambientes de aprendizagem on-line. O potencial destes ambientes, no entanto, pode não se concretizar por problemas decorrentes da separaçao física entre professor e alunos. Neste sentido, surgem soluçoes utilizando as técnicas de data mining. O objetivo deste artigo é analisar o uso destas técnicas em ambientes de aprendizagem on-line, uso este bastante mencionado na literatura, mas pouco documentado. As conclusões apontam que este uso ainda é restrito, mas os exemplos analisados mostram que a adoçao de técnicas de data mining pode ser bastante proveitosa.*

## 1. Introduçao

Uma das vantagens dos ambientes de aprendizagem *on-line* é a independência da presença física dos participantes no mesmo espaço geográfico, permitindo que os estudantes tenham a possibilidade de adquirir conhecimento sobre assunto de seu interesse em locais diferentes da sala de aula convencional. No entanto, tal independência pode ocasionar alguns problemas. Nestes ambientes, uma atividade difícil de ser realizada pelo professor é a análise das discussões em andamento, para verificar se o escopo da discussão é mantido e se todos os alunos participam dos debates. Problemas relacionados à coordenação das atividades de aprendizagem, à avaliação do desempenho do aluno e à falta de foco nas discussões que ocorrem nas ferramentas de comunicação, exigem o uso de soluções computacionais.

Seria interessante que algum suporte de software estivesse disponível visando observar as interações, extrair informações a partir dessas interações, realizar análises e transmitir os resultados ao professor. Neste sentido, o uso de técnicas de *data mining* (ou mineração de dados) tem sido proposto. *Data mining* utiliza técnicas estatísticas e de aprendizado de máquinas para construir modelos

capazes de prever o comportamento dos dados. Segundo (www1), o uso de *Data mining* para construção de um modelo traz como vantagens:

- Modelos são de fácil compreensão: pessoas sem conhecimento estatístico podem interpretar o modelo e compará-lo com suas próprias idéias. O usuário ganha mais conhecimento sobre o comportamento do cliente e pode usar esta informação para otimizar os processos dos negócios.

- Grandes bases de dados podem ser analisadas: grandes conjuntos de dados podem ser analisados com *Data mining*. Por exemplo, para cada cliente, pode-se ter centenas de atributos que contêm informações detalhadas. Bases de dados podem ser muito extensas e se pode querer minerar uma base de dados contendo de registros de clientes.

- *Data mining* descobre informações não esperadas: como muitos modelos diferentes são validados, alguns resultados inesperados podem surgir. Em diversos estudos, descobriu-se que combinações de fatores particulares tiveram resultados inesperados.

- Variáveis não necessitam de recodificação: *Data mining* lida tanto com variáveis numéricas quanto categóricas. Estas variáveis aparecem no modelo exatamente da mesma forma em que aparecem na base de dados.

- Modelos são precisos: os modelos obtidos por *Data mining* são validados por técnicas de estatística. As previsões feitas por estes modelos são precisas.

- Modelos são construídos rapidamente: *Data mining* permite gerar modelos atualizados em poucos minutos, ou poucas horas. A modelagem se torna mais fácil já que os modelos são testados, e apenas os melhores modelos são retornados aos usuários.

Considerando os problemas enfrentados por professores e alunos em ambientes *on-line* e o potencial das técnicas de *data mining*, o objetivo deste artigo é analisar o uso destas técnicas no acompanhamento dos alunos *on-line*. Para atingir seu objetivo, o artigo está organizado da seguinte forma: a seção 2 discute os problemas que podem ocorrer em ambientes *on-line*; a seção 3 apresenta

---

\* Artigo extraído do projeto de graduação “Técnicas de *Data mining* em Ambientes de Aprendizagem *On-line*” apresentado ao DICC/IME/UERJ pelos alunos Fabrcio R. Cunha e Rafael L. P. Araujo.

uma visão geral de *Data mining*, a seção 4 descreve como as técnicas de *Data mining* estão sendo empregadas para apoiar a aprendizagem *on-line* e a última seção oferece as conclusões do trabalho.

## 2. Problemas dos Ambientes de Aprendizagem *On-line*

Silva e Santos (2002) identificaram um conjunto de problemas referentes à aprendizagem *on-line*:

- Baixa interação - em ambientes EAD, é difícil para o professor analisar as discussões em andamento, verificando o escopo da discussão e incentivando a participação dos alunos.
- Falta de Interesse dos Alunos - a aprendizagem *on-line* permite que os cursos sejam assistidos por um número grande de alunos, com perfis de interesse e *background* diferenciados. Quanto mais esta potencialidade é introduzida, mais existe o risco do estudante não atingir seus objetivos acadêmicos, subutilizando o suporte oferecido.
- Evasão de Estudantes - pela separação geográfica e temporal entre os participantes dos cursos, há grande evasão de estudantes.
- Falta de Monitoramento/Coordenação - os alunos não realizam cooperativamente a tarefa proposta e/ou se envolvem pouco em sua realização porque falta a presença de um coordenador para guiá-los e dar conselhos ou a presença do coordenador não têm suporte efetivo no ambiente de aprendizagem cooperativa.
- Baixa Participação/Adesão dos Alunos - em ambientes de aprendizagem cooperativa, nem sempre os níveis de cooperação são os esperados, porque os alunos se engajam pouco nas atividades cooperativas, não assumem as responsabilidades exigidas e às vezes não entendem seu papel na execução da tarefa.
- Conflito entre os Participantes - durante as sessões de aprendizagem cooperativa, podem surgir conflitos entre os membros do grupo, acarretando problemas na execução das tarefas. Embora os conflitos sejam parte inerente do processo de cooperação, eles podem refletir desavenças pessoais e enfraquecer a coesão do grupo. É necessário que o ambiente de aprendizagem cooperativa ofereça formas de resolução de conflitos.
- Dificuldades na Representação do Conhecimento em Construção - em ambientes cooperativos, dependendo da atividade cooperativa desenvolvida por um grupo, pode ser fundamental que as pessoas tenham mecanismos formais, estruturados, para representar um conhecimento, ou questionar uma outra colocação, de forma que todos os participantes tenham oportunidade de entender o que se está querendo comunicar.

- Pouca Sistematização do Conhecimento - A meta de muitos ambientes de aprendizagem cooperativa é incentivar a cooperação como um valor social. Logo, as relações lógicas envolvidas no conhecimento que está sendo manipulado podem não se tornar claras para os alunos.

- Baixa Motivação pela Falta da Presença dos Participantes - os ambientes de aprendizagem cooperativa permitem que estudantes trabalhem juntos, compartilhando espaços virtuais, mas estes ambientes, contudo, não conseguem reproduzir as mudanças que ocorrem em uma situação de interação face a face. A percepção é um conceito relacionado a mecanismos que garantem que as pessoas compreendam ou tomem consciência do próprio processo e da interação entre todos os participantes no ambiente. Os elementos de percepção são essenciais para que os estudantes possam aprender e trabalhar em equipe.

Grande parte das soluções apontadas e implementadas para estes problemas adota agentes de software. Mas, recentemente, técnicas de *Data mining* foram introduzidas na modelagem das soluções.

## 3. Introdução a *Data mining*

A mineração de dados é a exploração e a análise, por meio automático ou semi-automático, de grandes quantidades de dados, a fim de descobrir padrões e regras significativos. Ela também é conhecida como *Descoberta de Conhecimento em Bases de Dados (Knowledge Discovery in Databases - KDD)*, e segundo Han & Kamper (2000) se refere à extração não trivial, implícita, desconhecida previamente e potencialmente útil de informações ou padrões de em grandes bases de dados.

Os principais objetivos da mineração de dados são descobrir relacionamentos entre dados e fornecer subsídios para que possa ser feita uma previsão de tendências futuras baseada no passado. Os resultados obtidos com a mineração de dados podem ser usados no Gerenciamento de Informação, Processamento de Pedidos de Informação, Tomada de Decisão, Controle de Processo e muitas outras aplicações.

KDD, como um processo, consiste de uma sequência interativa e iterativa das seguintes etapas: Seleção, Pré-processamento, Transformação, Garimpagem (Mineração) e Análise e Assimilação dos Resultados (Cabena *et al.* 1998):

- Seleção – selecionar ou segmentar dados de acordo com critérios definidos. Ex.: “todos alunos inscritos em determinado módulo” é um subconjunto de dados determinado.
- Pré-processamento – estágio de limpeza dos dados, onde informações julgadas desnecessárias são removidas. Ex.: sexo do aluno. Reconfiguração dos

dados para assegurar dados consistentes de identificação. Ex.: sexo = "F" ou "M" e sexo = "M" ou "H".

- Transformação – transformação dos dados em formatos utilizáveis. Ex.: rede neural  $\Rightarrow$  converter valor literal em valor numérico. Disponibilizar os dados de maneira usável e navegável.

- Mineração – a verdadeira extração dos padrões de comportamento dos dados.

- Análise e Assimilação dos Resultados – identificado os padrões pelo sistema, estes são interpretados em conhecimentos e validados, os quais darão suporte a tomada de decisões humanas. Ex.: tarefas de previsões ou classificações.

Como se trata de um processo iterativo, uma vez que o conhecimento descoberto é apresentado ao usuário, as medidas de validação podem ser melhoradas, a "mineração" em si pode ser refinada, novos dados podem ser selecionados, ou novas fontes de dados podem ser integradas, a fim de obter-se resultados diferentes ou mais apropriados.

Em princípio, *Data mining* não seria específico a um tipo de dado e poderia ser aplicável a qualquer repositório de dados. Contudo, algoritmos e abordagens podem diferir quando aplicados a diferentes tipos e estruturas de dados. Realmente, os desafios apresentados por diferentes tipos de dados podem variar significativamente. *Data mining* vem sendo usado e estudado em bancos de dados relacionais, bancos de dados orientados a objeto, *Data Warehouses*, bancos de dados transacionais, repositórios não-estruturados ou semi-estruturados tais como a Internet e bancos de dados avançados com conteúdo multimídia e geográfico. Atualmente, o foco maior está em bancos de dados relacionais e em *Data Warehouses*. Outros enfoques ainda precisam de mais desenvolvimento para aplicação em estruturas ou tipos de dados mais complexos.

Algoritmos de *Data mining* para bancos de dados relacionais podem ser mais versáteis do que algoritmos escritos para arquivos sem estruturação - simples arquivos em formato texto ou binário, como transações, dados série-temporal, etc. - já que podem tomar proveito da estrutura inerente a este tipo de bancos de dados. *Data mining* pode se aproveitar da SQL para seleção, transformação e consolidação de dados, mas vai além do que a SQL poderia prover, permitindo predição, comparação, detecção de padrões, etc.

*Data warehouse* é um repositório de dados coletados de múltiplas fontes, geralmente heterogêneas, e é planejado para ser usado como uma fonte de dados única ou unificada. *Data warehouse* oferece a possibilidade de se analisar dados de diferentes bases de dados e diferentes estruturas pré-existentes, sob uma ótica homogênea. A fim de facilitar a tomada de decisão e permitir

visões multidimensionais dos fatos representados, *Data Warehouses* são modelados através de uma estrutura de dados multidimensional.

As técnicas de mineração de dados podem ser aplicadas a tarefas como Classificação, Estimativa (ou Regressão), Associação, Segmentação e Sumarização. Essas tarefas são descritas a seguir:

- **Classificação** - A tarefa de classificação consiste em construir um modelo que possa ser aplicado a dados não classificados visando categorizá-los em classes. Um objeto é examinado e classificado de acordo com uma classe definida. A tarefa de classificação pode ser considerada uma tarefa mal definida, indeterminística, que é inevitável pelo fato de envolver predição. São exemplos de tarefas de classificação: classificar pedidos de créditos como de baixo, médio e alto risco; esclarecer pedidos de seguros fraudulentos; identificar a forma de tratamento na qual um paciente está mais propício a responder, baseando-se em classes de pacientes que respondem bem a determinado tipo de tratamento médico.

- **Estimativa (ou Regressão)** - A estimativa é usada para definir um valor para alguma variável contínua desconhecida como, por exemplo, receita, altura ou saldo de cartão de crédito. Ela lida com resultados contínuos, enquanto que a classificação lida com resultados discretos. Ela pode ser usada para executar uma tarefa de classificação, convencendo-se que diferentes faixas (intervalos) de valores contínuos correspondem a diferentes classes. "Regressão é aprender uma função que mapeia um item de dado para uma variável de predição real estimada" (Fayyad, 1996). Como exemplos de tarefas de estimativa tem-se: estimar o número de filhos em uma família; estimar a renda total de uma família; estimar o valor em tempo de vida de um cliente; estimar a probabilidade de um paciente morrer, baseando-se nos resultados de um conjunto de diagnósticos médicos; prever a demanda de um consumidor para um novo produto.

- **Associação** - Análise por associação é a descoberta do que se chama regras de associação. Ela descreve a frequência dos itens que aparecem juntos em bancos de dados baseada em dois conceitos: suporte e confiança. O primeiro quantifica o intervalo em que foi descoberta a associação e o segundo quantifica a probabilidade dos itens considerados aparecerem juntos em uma transação. As regras de associação descobertas são da forma  $P \rightarrow Q[s,c]$ , onde P e Q são as conjunções dos atributos considerados, s (de suporte) é a probabilidade de que P e Q apareçam juntos numa transação e c (de confiança) é a probabilidade de Q aparecer numa transação quando P está presente. Por exemplo, a regra de associação hipotética: *Curso(X,"BD")^Idade(X,"19-25")*

$\rightarrow$ *Curso(X,"Access")[s=2%,c=55%]* indicaria que

2% das transações consideradas são de alunos entre 19 e 25 anos que estariam participando do curso BD e Access ao mesmo tempo, e que há uma certeza de 55% que jovens dessa idade que participam do curso de BD também participam do curso de Access. “A tarefa de associação pode ser considerada uma tarefa bem definida, determinística e relativamente simples, que não envolve predição da mesma forma que a tarefa de classificação”.

- **Segmentação (ou Clustering)** - A segmentação é um processo de partição de uma população heterogênea em vários subgrupos ou *clusters* mais homogêneos. Na segmentação, não há classes predefinidas, os registros são agrupados de acordo com a semelhança, o que a diferencia da tarefa de classificação. Exemplos de segmentação: agrupar os clientes por região do país, agrupar clientes com comportamento de compra similar; agrupar seções de usuários Web para prever comportamento futuro de usuário.

- **Sumarização** - Segundo Fayyad (1996), a tarefa de sumarização envolve métodos para encontrar uma descrição compacta para um subconjunto de dados. Um simples exemplo desta tarefa poderia ser tabular o significado e desvios padrão para todos os itens de dados. A Sumarização engloba:

- **Caracterização de dados**, que é uma sumarização das propriedades gerais de objetos numa classe alvo, e produz regras de caracterização. Os dados relevantes para uma classe definida pelo usuário são normalmente disponibilizados através de uma consulta de banco de dados e processados por um módulo de sumarização, para extrair a essência dos dados em diferentes níveis de abstração. Por exemplo, pode-se querer caracterizar os clientes que regularmente alugam mais de 30 filmes por ano numa locadora; e

- **Discriminação dos dados** que produz regras de discriminação e são basicamente uma comparação de propriedades gerais de objetos entre duas classes, referenciadas como *tarje calas* e *contrastem calas*. Por exemplo, poderia se comparar as características gerais de consumidores que alugaram mais de 30 filmes ano passado com aqueles cuja cota não passou de 5. As técnicas usadas para discriminação são bem similares às técnicas de caracterização com a exceção que os resultados da discriminação incluem medidas comparativas.

Segundo Han e Kamper (2000), as técnicas de mineração de dados podem ser aplicadas a tarefas como Análise de exceções, ou seja, dados que não podem ser agrupados numa classe pré-estabelecida, e também Análise de evolução e desvio, que diz respeito ao estudo de dados que mudam com o tempo. Análise de exceções, também conhecidas como *outliers* ou surpresas, é a análise dos dados que tanto podem ser descartados em algumas

aplicações, propiciando eficiência, como podem revelar importantes conhecimentos em outros domínios. A análise de evolução modela tendências usando caracterização, comparação, classificação ou agrupamento de dados relacionados com o tempo.

A gama de informações coletadas atualmente é bastante extensa - de simples medidas numéricas e documentos de texto às mais complexas informações como dados geográficos, dados multimídia e documentos hipertextos. É comum encontrar em ambientes de educação *on-line*, que é o foco deste trabalho, análises gerenciais através de estatísticas sobre dados coletados referentes ao cumprimento de atividades planejadas e às interações dos alunos. Essas informações ficam armazenadas em bases de dados de material de controle, ou seja, uma base de dados (orientada a objetos ou relacional) elaborada para armazenar e gerenciar informações referentes a cursos, disciplinas, professores, alunos, matrícula, avaliações, acompanhamento de alunos, entre outras.

Para melhor explorar essas informações, ou seja, extrair a chamada essência da informação armazenada, propõe-se analisá-las com o auxílio de algoritmos de *Data mining* para encontrar padrões relativos ao comportamento dos alunos em cursos a distância, com o objetivo de obter um conjunto de regras que favoreçam o acompanhamento e a avaliação do aprendizado do aluno. Por exemplo, a aplicação de uma tarefa de *Data mining* chamada de associação sobre os dados armazenados ao longo do tempo num banco de dados poderia considerar as atividades planejadas para o aluno e as atividades que o aluno efetivamente cumpriu, indicando, hipoteticamente, que, por exemplo, de 80% dos alunos que cumprem todas as atividades, um percentual de 60% é aprovado [Silva, 2001].

Outro exemplo seria classificar um aluno como desistente em potencial do curso, com base em seu padrão de comportamento e no padrão obtido pelos dados históricos de outros alunos.

O uso de técnicas de mineração de dados em ambientes de aprendizagem *on-line* pode ser de fundamental importância, pois podemos descobrir vários padrões e regras que estavam escondidos em volumosos dados, através de consultas *ad hoc*, e que seriam dificilmente visualizadas de outra forma. Na próxima seção, apresentamos a utilização de *Data mining* em ambientes de aprendizagem *on-line*.

#### 4. *Data mining* em Ambientes de Educação *On-line*

Como mencionado, diversos problemas podem ocorrer no processo educacional mediado pela Web. A maior parte das soluções para problemas descritos adota agentes de software (Silva & Santos, 2001; Santos, 2003). Mas alguns ambientes Web

para suporte à aprendizagem *on-line* estão empregando os conceitos de *Data mining*.

Acompanhar o progresso do aluno é uma tarefa difícil. De um modo geral, o professor utiliza instrumentos formais (provas, listas de exercício) para este fim. Mas a utilização de instrumentos informais, como a observação do comportamento do aluno, sua participação em atividades, debates, pode ser valiosa. Em atividades presenciais, nem sempre o professor consegue acompanhar o progresso do aluno com instrumentos informais. Em ambientes *on-line*, este acompanhamento pode ser mais facilmente realizado e para tanto, tem sido proposta a análise das mensagens trocadas pelos alunos com a mediação de ferramentas de comunicação. É neste sentido que as técnicas estão sendo empregadas. Examinamos, a seguir, exemplos disponíveis.

#### 4.1. Exemplos de *Data mining* em Ambientes de Educação *On-line*

Há alguns trabalhos visando facilitar o acompanhamento e a análise dos dados gerados pelas ações dos alunos em ambientes *on-line*. Algumas pesquisas empregam agentes de *software* que atuam filtrando e analisando as participações dos alunos por meio dos registros das interações ou acompanhando, analisando e gerando *feedback*. Também são encontradas pesquisas que envolvem o registro das interações dos alunos e aplicação de técnicas de *data mining* para a extração e inferência de padrões de comportamento.

São ainda poucos os ambientes de aprendizagem *on-line* que adotam técnicas de *data mining*, mas alguns exemplos disponíveis são bastante interessantes. Nas próximas seções, descrevemos três destes exemplos.

##### 4.1.1. Midas-Poeta

Midas-Poeta é sistema de apoio à decisão para o sistema Portfolio-Tutor e tem como objetivo oferecer um sistema tutor acoplado a um portfolio eletrônico (Nascimento, Rodrigues & Schiel, 2002).

O sistema tem duas camadas: a *camada portfolio* e a *camada tutor*. A camada portfolio implementa uma forma de avaliação auxiliando o professor no acompanhamento alternativo dos alunos. O sistema faz uso de um portfolio eletrônico no qual são depositados os artefatos gerados pelo aluno com as respectivas avaliações. O julgamento é realizado por itens subjetivos, como capacidade de síntese, liderança, criatividade entre outros, e também por itens somativos, onde uma nota é atribuída às atividades do aluno. A camada tutor tem como objetivo auxiliar o aluno em seção de aprendizagem visando classificá-lo em níveis de aprendizagem segundo o desenvolvimento de suas habilidades ao

longo do tempo. A base de dados do Portfolio-Tutor é relacional e foi implementada no SGBD Microsoft SQL Server 7.0.

O sistema é composto de quatro partes: usuários, estrutura curricular, portfolio e tutor. *Usuários*: armazenam as informações sobre os alunos, professores e as interações do aluno com os outros alunos através de trocas de mensagens de e-mails. Em seu modelo lógico, destacam-se:

- a tabela *ptuAluno*, que armazena informações demográficas (sexo, idade e região de procedência) e comportamentais (situação no curso) dos alunos; e
- a tabela *ptuEmail*, que representa as mensagens postadas pelo aluno durante uma disciplina.

A *Estrutura Curricular* armazena informações sobre disciplinas, cursos, matrícula em disciplinas. *Portfolio* armazenam os artefatos produzidos pelos alunos e os resultados dos itens de julgamento subjetivos feitos pelo professor. As tabelas *ptuUnidade* e *ptuAtividade*, que são as unidades planejadas em cada disciplina e as atividades realizadas pelo aluno em cada unidade, respectivamente.

Para compor o portfolio do aluno, cada atividade se refere a um artefato (*ptuArtefato*), que é avaliado segundo os itens de avaliação padronizados pela instituição de ensino (*ptuItemAvaliação*). Para um artefato, porém, podem existir diversas atividades associadas. O professor julga as atividades atribuindo notas aos itens de avaliação (*ptuAvaliacaoItemAvaliacao*). Para cada atividade realizada pelo aluno, são entregues os documentos que comprovam a sua realização. São exemplos de documentos: as respostas de um exercício, um resumo, um desenho, uma prova, enfim, qualquer artefato que represente a habilidade que o aluno adquiriu.

*Tutor* representa as ações realizadas pelo tutor para produzir uma seção de ensino virtual e as interações dos alunos com os materiais didáticos.

Apesar de ser um sistema que promove a aprendizagem *on-line* de forma transparente e efetiva, o Portfolio-Tutor apresenta a limitação de não proporcionar um módulo de análise que dê suporte a decisões estratégicas de aprendizagem. Com a utilização do portfolio, uma grande quantidade de dados fica disponível e estes dados podem revelar informações úteis (padrões, tendências, associações, etc) que favoreçam o acompanhamento do aprendizado dos alunos.

Para contemplar os aspectos alternativos do acompanhamento, são necessários dados sobre os itens de julgamento da avaliação autêntica. Sua base de dados prevê os dados necessários ao acompanhamento alternativo através da entidade *ptuAvaliacaoItemAvaliacao*, que armazenam os resultados dos itens de avaliação para cada artefato produzido pelo aluno.

Para acompanhar o processo de aprendizagem do aluno, o sistema Midas-Poeta coleta uma coleção de informações sobre o estudante que podem ser analisadas, e padrões comportamentais ou características interessantes dos alunos podem ser descobertos e usados como parâmetros para outros alunos.

Para tal, a estratégia deve contemplar as etapas de KDD, atender a mineração de dados interativa, onde o usuário, no momento da consulta possa selecionar os dados de interesse, selecionar a tarefa de mineração, informar medidas de interesse, capazes de filtrar as regras produzidas e visualizar e avaliar os resultados; dispor dados para o acompanhamento informal (interações aluno-pessoa e interações aluno-material didático); dispor dados para o acompanhamento alternativo (itens de portfólio); dispor dados sobre as avaliações, dados demográficos e comportamentais; e por fim, possibilitar o cruzamento destes dados coletados. Desta forma, uma grande base de dados pode ser produzida e utilizada para análises prospectivas.

A estratégia adotada em Midas-Poeta para apoiar decisões pedagógicas no acompanhamento do aprendizado é dividida em duas fases: a fase de planejamento e a fase de análise de dados.

Na fase de planejamento, definem-se os usuários do Sistema de Apoio a Decisão - SAD, a quais dados os mesmos terão acesso, que condições para a definição das informações de interesse o sistema de aprendizagem dispõe e que tarefas de mineração deverão estar disponíveis no SAD. Como produto da fase de planejamento, se tem um *Modelo Analítico de Dados - MAD*, que é o conjunto de todas as informações especificadas nos requisitos de dados da estratégia, na qual o usuário tem a possibilidade de navegar e selecionar os dados de interesse para uma consulta em questão. Sendo assim, o usuário usará o MAD para escolher os atributos que serão utilizados para análises prospectivas.

A fase de análise de dados é dividida em três etapas: etapa de coleta de dados, etapa de mineração de dados e etapa de visualização e incorporação dos resultados. Na etapa de coleta de dados, o usuário selecionará os atributos de interesse disponíveis no MAD, e necessários para uma mineração em questão. Na etapa de mineração de dados, o usuário selecionará uma tarefa de mineração prevista na fase de planejamento e assim produzirá o padrão de conhecimento que ele desejar.

Na etapa visualização e incorporação dos resultados, os resultados da mineração são visualizados e o usuário terá a possibilidade de decidir se o conhecimento produzido é útil ou não.

Para finalizar a definição das informações de interesse, é necessário definir os dados demográficos, os dados de avaliação e dados comportamentais do aluno. Com relação aos dados

demográficos e comportamentais, o Portfólio-Tutor apresenta um cadastro de alunos que armazena dados como idade, sexo, região de procedência e a situação do aluno no curso (desistente, trancado, cursando) e está disponível na tabela *ptuAluno*. Já para os dados sobre avaliação, o Portfólio-Tutor apresenta uma tabela que armazena as atividades realizadas pelo aluno, chamada *ptuAtividade*, e uma tabela que armazena o desempenho em cada atividade, chamada *ptuAtividadeAluno*.

De acordo com Nascimento, Rodrigues & Schiel (2002), o sistema Midas-Poeta foi desenvolvido baseado na arquitetura em três camadas, onde a interface, a lógica de negócio e a base de dados são independentes uma da outra. No tocante à implementação do sistema, foram utilizadas as linguagens Java e HTML/JSP para programar respectivamente a lógica de negócio e a montagem das interfaces Web. Para a implementação do modelo analítico de dados, utilizamos o SGBD SQL Server 7.0 e adotamos o mecanismo de visões.

Para atender a fase de análise de dados da estratégia de acompanhamento e o requisito de mineração de dados interativa, utilizamos uma interface para coleta de dados de interesse - *etapa de coleta de dados*, uma interface para seleção da tarefa de mineração - *etapa de mineração de dados*, e uma interface para apresentação dos resultados - *etapa de visualização e incorporação dos resultados*.

Com as interfaces do sistema, o usuário tem a possibilidade de selecionar os dados para uma consulta, a tarefa de mineração e visualizar os resultados. Caso o padrão produzido não tenha sido satisfatório, ele poderá voltar no processo e recomencá-lo.

A etapa de coleta de dados produz comandos *SQL* para realizar as junções entre as visões e atributos selecionadas pelo usuário gerando um *resultset* preparado para a mineração. Tal resultado deve conter apenas os atributos que foram selecionados pelo usuário podendo ser algum dos atributos sublinhados

#### **4.1.2. Ferramenta para Auxílio à Avaliação da Relevância das Mensagens e de Participações dos Alunos em Fóruns de Discussão** (Silva, Seno e Vieira, 2001).

De acordo com Silva, Seno e Vieira (2001), para acompanhar o processo de ensino-aprendizagem *on-line* é necessário definir um conjunto de informações que se julgue relevante ao acompanhamento. Então, é preciso estipular quais informações sobre o aluno e suas atividades são parâmetros para delinear o seu perfil e para medir o quanto ele aprendeu. Embora não haja um conjunto padrão de requisitos para acompanhar o aprendizado de um aluno, pode-se notar que há informações semanticamente

diferentes. Os autores propõem três etapas para orientar a implementação do acompanhamento do aprendizado em ambientes de EAD. Não há uma ordem cronológica na execução das etapas 1 e 2, e também não há obrigatoriedade na implementação das mesmas, pode-se implementar apenas uma das duas. Já a etapa 3 depende de qual das duas outras etapas foi implementada.

### **Etapas 1 – Coleta do Log e Rastreamento de Comunicação**

Essa etapa é responsável pela coleta e armazenamento das informações sobre o rastreamento das ações e comunicações do aluno. Nessa etapa, são armazenadas as interações do aluno com o ambiente e com os membros do curso. Classificam-se os tipos de interação que um aluno pode fazer:

- Aluno com outros integrantes do curso (professor, monitor ou outro aluno) sobre as quais é interessante saber, por exemplo, o assunto da mensagem e que mecanismo foi usado para enviá-la;
- Aluno com o material didático (páginas de conteúdo, provas, exercícios, sobre as quais é interessante saber, por exemplo, qual o tempo gasto nessa interação, se a interação foi de *download* ou *upload*, a qual disciplina pertence o material, qual *link* foi usado para chegar ao material, entre outros.

### **Etapas 2 – Reunião dos Resultados das Atividades Previstas**

As informações armazenadas nessa etapa são capazes de expressar o desempenho do aluno em termos das atividades propostas pelo professor. Tais informações são fortemente dependentes das atividades propostas para o curso, e da forma de validação escolhida para as mesmas, ou seja, do critério usado pelo professor para decidir se as atividades foram cumpridas ou não pelos alunos.

### **Etapas 3 – Análise dos Dados e Tomada de decisão**

A função dessa etapa é analisar os dados colhidos nas etapas 1 e 2, gerando ações para tratar as situações encontradas. As ações podem tanto estar a cargo do professor, cabendo a essa etapa somente informar ao professor o resultado das análises, quanto serem automáticas, acionando um mecanismo programado do sistema.

Após identificar os fatores que devem ser armazenados para acompanhar o aprendizado do aluno, é necessário que o professor analise essas informações. Os ambientes de EAD não oferecem recursos para apoiar análises mais sofisticadas sobre esses dados.

Silva, Seno e Vieira (2001) propõem que sejam usadas técnicas de *Data mining* para auxiliar o professor no acompanhamento do aluno. A aplicação de tarefas de *Data mining* sobre os dados

armazenados ao longo do tempo, relativos ao acompanhamento do aluno, poderá encontrar padrões de comportamento que estejam ligados ao processo de aprendizagem do aluno. Esses padrões podem, então, ser usados como parâmetros para outros alunos. Cabe ao professor decidir sobre quais aspectos aplica a mineração, o que acrescenta, ao acompanhamento, mais flexibilidade. Cada tarefa de *Data mining* produz um tipo diferente de padrão. Com base em dados hipotéticos, algumas tarefas estão exemplificadas abaixo, bem como os padrões que elas podem extrair.

#### **Tarefa de Classificação**

A classificação de dados é o processo de encontrar propriedades comuns entre um conjunto de objetos dentro de diferentes classes. Para exemplificar essa tarefa, e também as próximas, será utilizado parte de um banco de dados relacional de controle de oferta de cursos à distância.

A relação Interação Chat armazena o número de mensagens enviadas pelo aluno em uma sessão de Chat. A relação Aproveitamento contém dados sobre as questões o aluno acertou ou errou em uma avaliação formal e a nota que o aluno recebeu em cada questão. A relação Dedicção guarda os horários de acesso do aluno ao ambiente de aprendizagem. A relação Aluno tem informações pessoais do aluno e se ele desistiu do curso.

Para cada tarefa de *Data mining* a ser executada é necessário uma preparação dos dados: em algumas situações fazer junções entre tabelas e totalizar valores em colunas numéricas.

#### **Associação**

Podemos encontrar regras de associação entre atividades de um curso. Através de dados que se referem ao cumprimento das atividades de um curso pelos alunos podemos estabelecer dados estatísticos que nos fornecerão percentagens referentes à associação de atividades distintas.

#### **Agrupamento**

Aplicada ao acompanhamento da aprendizagem em EAD, podemos agrupar informações de acordo com algumas preferências: faixa etária, grau de instrução, conhecimento prévio do assunto da disciplina, e com isso gerar informações sobre o aproveitamento do aluno por determinado grau de instrução, determinada faixa etária.

#### **Regressão**

O uso dessa tarefa no acompanhamento do aprendizado do aluno pode prever o aproveitamento total do aluno considerando o tempo de permanência do mesmo no site e os dados históricos sobre a permanência de outros alunos. A partir desses dados podemos estabelecer uma relação

entre tempo de permanência no site e aproveitamento do aluno.

Na maioria dos ambientes de EAD que acompanham o aprendizado do aluno a exposição dos dados relativos a esse acompanhamento é feita de forma padronizada sem que o professor possa decidir quais aspectos lhe interessa ou quais informações deseja confrontar.

Com a utilização de *Data mining*, o professor ganha flexibilidade, pois pode confrontar os dados do acompanhamento pelo fato de poder selecionar os atributos do banco que lhe interessa e analisar a influência de alguns atributos sobre outros.

Os autores propõem um banco de dados para o acompanhamento do aprendizado. Tendo em vista que o nível de conhecimento, o grau de interesse e o fator tempo variam de aluno para aluno em um curso a distância, é desejável que o ambiente permita que o caminho ou ordem de execução do que foi planejado para um curso possa ser executado diferentemente de aluno para aluno e que a autorização para acessar o material didático do curso seja dinâmica de acordo com a ordem de execução escolhida pelo aluno. Com o objetivo de não impor obstáculos ao aprendizado, foi proposto um mecanismo para planejamento das atividades que respeita o ritmo de estudo de cada aluno. Durante a realização do curso, num determinado momento, cada aluno pode se encontrar em situações diferentes de aprendizado, dificultando o acompanhamento por parte do professor.

Assim, é mais que justificado que haja parâmetros para que se possa contar com um mecanismo para a orientação automática ao aluno. Esses parâmetros podem ser fornecidos por tarefas de *Data mining* aplicadas aos dados relativos aos cursos já realizados. Tais tarefas de *Data mining* podem ser aplicadas sobre os dados colhidos de um banco de dados para controle de oferta de cursos a distância, que apóia o funcionamento desse ambiente, onde ficam armazenadas as informações sobre os atores do processo de ensino (aluno, professor, monitor e administrador) e sobre o processo de ensino em si (planejamento e execução das atividades). Esse banco de dados está em constante evolução, uma vez que novas classes e atributos podem ser acrescentados em função da adição de novas atividades e de requisitos para as mesmas.

Esse banco de dados contém tanto informações pessoais desses atores quanto das atividades que os mesmos desenvolvem no curso. Por exemplo, sobre o aluno há informações pessoais e também dados das mensagens que envia para os demais participantes do curso, da execução das atividades planejadas para ele, da sua navegação nas páginas do curso, etc.

As informações da etapa 1, isto é, de rastreamento do aluno, estão armazenadas nas classes InteraçãoPessoal, InteraçãoPública e InteraçãoMaterialDidático. Essas classes abrangem os tipos de interação que o aluno pode realizar com o ambiente de aprendizagem.

Na classe InteraçãoPessoal são armazenados os dados sobre as mensagens particulares que o aluno envia para uma pessoa (Professor, Monitor ou outro Aluno) e também as mensagens que as pessoas enviam para o aluno. De cada interação desse tipo é necessário armazenar quem enviou a mensagem, qual aluno a recebeu, qual é o assunto da mesma, qual o mecanismo (*Chat*, *e-mail*, lista, etc) foi usado para enviá-la, qual a mensagem anterior a que ela se refere, qual a data em que a interação ocorreu, se ela é uma resposta à outra interação. Na classe InteraçãoPública estão os dados relativos às mensagens enviadas por uma pessoa para mais de um ator ao mesmo tempo, através de um mecanismo de envio de mensagens público, como lista de discussões ou *Chat*. Para tais interações são armazenados praticamente os mesmos atributos da InteraçãoPessoal.

Na classe InteraçãoMaterialDidático são armazenados os dados de cada acesso do aluno ao material didático de uma disciplina. Sobre esses acessos é interessante saber em qual página está o *link* de acesso à página requerida (LinkAcesso), que página é essa (EndereçoPágina), a hora em que a página foi aberta e o momento em que foi descarregada, a Disciplina à qual a página pertence, se a interação foi de *download* ou *upload*, entre outros.

As informações da etapa 2, referentes ao cumprimento das atividades previstas para uma turma de alunos, estão armazenadas na classe Cumprimento e dependem da checagem dos requisitos definidos pelo professor para cada atividade. Toda atividade prevista pelo professor para uma turma deve ser cumprida por cada aluno dessa turma, para cada um deles é registrado seu aproveitamento. Esse aproveitamento pode estar vinculado ao atendimento dos requisitos de realização da atividade (RequisitoRealizado). Cada atividade tem seus próprios requisitos (RequisitoChat, RequisitoLeitura, RequisitoTrabalho, RequisitoQuestionário, RequisitoOutraAtividade), que podem também estar ligados a resolução de questões sobre a atividade. Essas questões podem ser dissertativas, de teste ou V/F.

Cada classe tem atributos que apóiam análises gerenciais; maiores informações sobre os mesmos podem ser encontradas em [8] e [10]. Com as informações armazenadas nesse banco de dados pode-se implementar a etapa 3 para acompanhamento, ou seja, a etapa de análise dos

dados e tomada de decisão. Essa implementação é justamente o diferencial desta pesquisa, pois as informações armazenadas ao longo do tempo possibilitam a utilização de técnicas mais refinadas de análise, como as descritas na seção 3.2. Está em fase de especificação uma ferramenta para apoiar essa etapa do acompanhamento. Tal ferramenta tem duas funções básicas: pré-processar os dados do banco de dados de controle de oferta de cursos a distância, selecionando as informações relevantes, preparando-as para a mineração e armazenando-as num *Data Warehouse*; e obter padrões sobre o comportamento dos alunos, aplicando algoritmos de *Data mining* sobre os dados do *Data Warehouse*.

Na maioria dos ambientes de EAD que, de alguma forma, acompanham o aprendizado do aluno, a exposição dos dados relativos a esse acompanhamento é feita de forma padronizada, ou seja, sem que o professor possa decidir quais aspectos lhe interessam ou quais informações deseja confrontar.

O objetivo do mecanismo proposto por Silva, Seno e Vieira (2001) é minimizar essas dificuldades. Sendo *Data mining* uma forma de apoiar a tomada de decisão, sua utilização no processo de acompanhamento do aprendizado auxilia o professor na análise do comportamento dos alunos para que o mesmo possa, mais facilmente, orientá-lo durante o curso. O ponto forte deste mecanismo é o ganho de flexibilidade que o professor tem para escolher e confrontar os dados do acompanhamento, pois ele pode selecionar os atributos do banco de dados que lhe interessar e analisar a influência de alguns atributos sobre outros.

Silva, Seno e Vieira (2001) implementaram uma ferramenta para auxílio à avaliação da relevância das mensagens e de participações dos alunos em fóruns de discussão. Com a estrutura proposta, é possível analisar não somente estatísticas como também identificar padrões de comportamento, através de tarefas *Data mining*, ganhando grande flexibilidade no processo de avaliação subjetiva.

Na 1ª etapa as informações coletadas dizem respeito às interações do aluno com o ambiente e do aluno com outros alunos e quais meios foram utilizados para isto. As interações possíveis em um ambiente EAD são: aluno-aluno, aluno-professor, aluno-tutor, aluno-material didático (páginas de conteúdo, testes, exercícios, etc.).

Na 2ª etapa as informações coletadas dizem respeito à performance do aluno em termos das atividades propostas pelo professor. As informações coletadas nesta etapa dependem das atividades propostas para o curso e dos modelos de validação, regras impostas pelo professor, escolhidos para tais atividades. Por exemplo, cada atividade proposta pode produzir um resultado como uma nota num teste, a participação ou não em um seminário, o

número de mensagens enviadas numa conferência, etc.

Na 3ª etapa acontece a análise propriamente dita dos dados coletados nas etapas anteriores. Acompanhamento contínuo significa acompanhar os passos dos alunos durante o curso, guiando-o, sempre quando necessário, através do caminho aprendido. Neste sentido, as etapas 1 e 2 trilham as atividades dos alunos enquanto a 3ª etapa fornece a base para que o professor possa ajudar o estudante a seguir o melhor caminho. Baseadas nas análises geradas nesta etapa, ações podem ser tomadas por professores ou pelo próprio sistema.

A performance do estudante é refletida em várias situações, e o professor pode analisar cada situação em combinação com outras características tais como número de vezes que o aluno acessou materiais didáticos e participou de conferências. O professor então toma a decisão que achar mais apropriada, baseada nesta análise.

Todas interações são armazenadas em ordem cronológica, com cada interação associada a uma data. Com as informações coletadas de acordo com esse modelo, é possível responder perguntas do tipo:

- Quanto tempo um aluno gastou no curso ou num determinado assunto?
- Em que ordem o aluno acessou as páginas do curso?
- O nível de interesse do aluno pelo curso mudou?
- Que assuntos o aluno procurou discutir?

Para a implementação da 2ª etapa não foi apresentado um modelo de estrutura, pois as informações coletadas nesta etapa, dependem do método que o professor escolher para validar as atividades previstas. Um ambiente padrão possui atividades comuns como testes, exercícios, trabalhos de leitura e escrita, debates, conferências, etc. As informações coletadas nesta etapa podem responder questões tais como:

- Com que frequência um aluno se engaja numa atividade?
- Qual a nota de um aluno em uma atividade específica?
- Quais os pontos bons e ruins de um aluno?
- Qual a nota média de um aluno numa determinada atividade?

Tendo guardado os dados relativos às etapas 1 e 2, na 3ª etapa se dará a análise não estática desses dados com auxílio de tarefas de *Data mining*, identificando padrões de comportamento dos alunos. Por exemplo, uma tarefa *Data mining* chamada associação poderia, levando-se em conta atividades planejadas e atividades realmente executadas pelos alunos, indicar que de 80% dos alunos que cumprem todas as atividades 60% são aprovados no curso.

#### 4.1.3. Uso de *Data Warehouses* e *Data mining* na avaliação de estudantes em cursos à distância

Uma aplicação interessante das técnicas de data mining utiliza *Data Warehouses* em ambientes *on-line*, com o objetivo de aproveitando-se de sua estrutura própria destes ambientes, auxiliar o processo de avaliação dos alunos. A aplicação baseia-se na premissa de que todo ambiente *on-line* mantém informações sobre as atividades realizadas pelos alunos. Analisar esses históricos de informações pode revelar como a maneira que o aluno conduz seus estudos pode influenciar na integralização do curso. Cinco mecanismos de suporte são identificados na maioria desses ambientes: rastreamento das ações do aluno, redirecionamento através de testes, registros de listas, registros de fóruns e registro de chats.

Para os autores, a maioria dos ambientes disponíveis contém um pequeno conjunto de informações sobre o caminho percorrido pelo aluno durante o curso. Mas, não há um padrão para esse conjunto de informações, porém há um consenso sobre dois tipos de informações que servem para guiar a tarefa de acompanhar o aproveitamento de alunos em ambientes EAD:

- Informações sobre a interação do aluno com outros participantes do ambiente visando analisar a influência da interação (aluno-pessoa e aluno-material) com o aproveitamento.
- Informações sobre as atividades do aluno no curso.

Uma tabela *PersonalInteraction* especializa em quatro tipos de interações diferentes. Ela guarda informações sobre como o aluno interage com o ambiente, mas não guarda a informação trocada propriamente dita. Por exemplo, não guarda a mensagem enviada de um aluno para outro numa sessão de Chat.

Além das informações sobre atividades e interações pessoais, este *Data Warehouse* contém as seguintes informações:

- Interação do aluno com o material didático, envolvendo atributos como duração do acesso, link do material e tipo de acesso (upload ou download).
- Os testes que o aluno realizou com os atributos nota, nº de respostas incorretas, etc.
- E se o estudante passou nos testes de conclusão de uma disciplina, com os atributos abandonou? passou?

O *Data Warehouse* apresenta a possibilidade de análises variadas por combinar medidas e dimensões, por exemplo:

- Analisar se há relação entre o aproveitamento do aluno, suas interações pessoais e interações com materiais didáticos (envolvendo os fatos *Test*, *PersonalInteraction* e *StudentMaterialInteraction*);

- Verificar a influência de fatores tais como tipo de comunicação e interação com materiais didáticos (envolvendo os fatos *PersonalInteraction* e *StudentMaterialInteraction*);

- Descobrir se o tipo de conexão influencia no número de vezes que o aluno acessa o ambiente (envolvendo a dimensão *Student* e o fato *StudentMaterialInteraction*);

- Descobrir atividades que sejam mais eficazes para determinados cursos, faixas etárias, graus de escolaridade, etc. (envolvendo as dimensões *Course* e *Student* e o fato *Activity*).

Estas análises podem ser realizadas usando técnicas de *Data mining* um ambiente de KDD (knowledge Discover in Databases).

Nesta seção do trabalho, procuramos identificar como as técnicas de *data mining* estão sendo usadas no acompanhamento da aprendizagem de alunos nos cursos à distância na Web. Da análise realizada, verificamos que este uso ainda é muito restrito e após uma busca exaustiva em fontes de consulta disponíveis, localizamos somente três propostas de utilização de data mining em ambientes educacionais. Mas, nos três exemplos analisados, verificamos que o principal papel desempenhado por estas técnicas é o de auxiliar o professor no monitoramento dos debates que ocorrem nos fóruns de discussão e nas sessões de *Chat*, visando acompanhar o processo de aprendizagem e fornecer subsídios para a avaliação dos alunos.

A quantidade de mensagens trocadas pelos alunos nos chats e nos fóruns pode tornar muito trabalhosa a análise dos conteúdos destas mensagens. Neste sentido, a utilização das técnicas de *data mining* seria bastante adequada, já que a mineração de dados se refere, especificamente, à extração não trivial, implícita, desconhecida previamente de informações ou padrões de dados úteis em grandes bases de dados (Am & Kandir, 2000).

## 5. Conclusões

Nosso objetivo neste artigo foi analisar o uso de técnicas de *data mining* para auxiliar o acompanhamento do aluno em ambientes de aprendizagem *on-line*. Vimos que a modelagem e o desenvolvimento de técnicas de *data mining* são tarefas não triviais. *Data mining* é um processo que consiste de uma seqüência interativa e iterativa (Cabana *ect al.* 1998). Este processo começa com o entendimento e definição do domínio da aplicação e dos objetivos finais a serem atingidos. Para isso, é necessária a presença de um analista da área de aplicação, que possua todo o entendimento sobre o domínio em questão. Seguindo o processo, temos a etapa de Seleção, onde são identificados as bases de

dados e quais atributos e exemplos de dados devem ser extraídos para a fase de Garimpagem.

A fase de Pré-processamento destina-se a melhorar a qualidade dos dados e conseqüentemente os resultados da mineração, pois os dados selecionados na etapa anterior podem estar inconsistentes e incompletos.

A etapa de Transformação de Dados visa adequar os dados selecionados e pré-processados de acordo com a entrada específica de cada algoritmos de mineração de dados.

Na fase de Garimpagem são recebidos os dados preparados pelas fases anteriores, e é aplicado um algoritmo de mineração de dados que vai produzir um modelo de conhecimento. Tal modelo pode utilizar regras de associação, classificadores, árvores de decisão, etc. Esta etapa é particularmente complexa e trabalhosa.

Por fim, a etapa de Análise e Assimilação dos Resultados, irá gerar relatórios, modelos ou representações dos padrões descobertos, que passam ser interpretados e validados pelo especialista do domínio, para que o modelo produzido seja relevante para o problema em questão.

Revisamos os conceitos básicos de *data mining* com o propósito de analisar como estes conceitos estão sendo usados nos ambientes educacionais. Midas-Poeta, por exemplo, usa técnicas de *data mining* para gerar informações úteis sobre o processo de aprendizagem (Nascimento, Rodrigues & Schiel, 2002). No sistema, uma coleção de informações sobre o estudante e padrões comportamentais ou características interessantes dos alunos podem ser descobertos e usados como parâmetros para outros alunos. A estratégia empregada contempla as etapas de *data mining*, onde o professor seleciona os dados de interesse, seleciona a tarefa de mineração, informa as medidas de interesse, capazes de filtrar as regras produzidas e visualiza e avalia os resultados.

As tarefas de mineração previstas para o sistema Midas-Poeta são as tarefas de associação e classificação. Os autores acreditam que operações preditivas e classificatórias se adequam melhor à tomada de decisão na área educacional. Como visto, a classificação é o processo de prever a classe de um objeto segundo certos atributos de predição. Já as regras de associação são aquelas que correlacionam itens no formato *atributo=valor* sem que nenhuma hipótese seja pré-estabelecida.

Midas-Poeta usa uma estratégia de acompanhamento do aprendizado, na qual, a partir de uma base de acompanhamento que contempla dados informais, alternativos, demográficos e comportamentais dos alunos, são utilizadas técnicas de mineração de dados para extrair padrões úteis e que favoreçam o processo de aprendizagem. O sistema é bastante interessante, mas convém

ressaltar que a parte relativa à mineração de dados não está implementada.

Silva, Seno e Vieira (2001) entendem que cada tarefa de *data mining* produz um tipo diferente de padrão e os diferentes padrões a serem descobertos podem requer o emprego de uma ou mais tarefas de classificação, associação, agrupamento e regressão. Estes autores implementaram uma ferramenta para auxílio à avaliação da relevância das mensagens e de participações dos alunos em fóruns de discussão. Com a estrutura proposta, é possível analisar não somente estatísticas como também identificar padrões de comportamento, através de tarefas *data mining*.

As informações sobre as interações dos alunos são coletadas, bem como as informações sobre o desempenho do aluno nas atividades propostas pelo professor. A seguir, os dados coletados nas etapas anteriores são analisados. As tarefas de *data mining* utilizadas visam responder a consultas do tipo: Quanto tempo um aluno gastou no curso ou num determinado assunto? Em que ordem o aluno acessou as páginas do curso? O nível de interesse do aluno pelo curso mudou? Que assuntos o aluno procurou discutir? Com que frequência um aluno se engaja numa atividade? Qual a nota de um aluno em uma atividade específica? Quais os pontos bons e ruins de um aluno? Qual a nota média de um aluno numa determinada atividade?

Um terceiro exemplo examinado usa *Data Warehouse* e *data mining* na avaliação de estudantes em cursos *on-line*. A idéia é desenvolver um modelo de *Data Warehouse* capaz de fornecer suporte para a extração de informações sobre a interação do aluno com outros participantes do ambiente e sobre as atividades do aluno no curso. Além das informações sobre atividades e interações pessoais, o *Data Warehouse* proposto contém informações sobre a Interação do aluno com o material didático, os testes realizados pelo aluno, nota, nº de respostas incorretas, etc.

O aspecto mais interessante e promissor do modelo de *Data Warehouse* apresentado é a possibilidade de combinar dimensões, para analisar a eventual relação entre o aproveitamento do aluno, suas interações pessoais e interações com materiais didáticos e para descobrir atividades eficazes para determinados cursos, faixas etárias e graus de escolaridade.

Analisando os trabalhos disponíveis nas áreas de aprendizagem *on-line* e de *data mining*, concluímos que as tarefas de *data mining* potencialmente mais úteis para auxiliar o professor no acompanhamento da aprendizagem e na avaliação dos alunos, se baseando nas mensagens trocadas nos fóruns e chats, seriam:

- Tarefas de classificação – podemos classificar os alunos em categorias por nível de

participação e por contribuição na construção da solução. Além dos dados analisados pela avaliação manual do professor (incluindo a avaliação das mensagens, os autores, o grupo), poderíamos classificar se o aluno teve uma participação “alta”, “média” ou “baixa”. Mas precisaríamos da mineração de dados para procurar por padrões que indiquem a qualidade da participação.

- As tarefas de regressão – com elas podemos estimar o número de mensagens que um aluno vai enviar e avaliar durante o curso. Essa previsão seria interessante para o professor saber qual seria a contribuição esperada de cada aluno no fórum ou nas sessões de Chat.

- As tarefas de associação – com elas podemos associar os tópicos e termos que mais se interessam aos alunos, observando o seu comportamento no sistema para montar uma lista de assuntos de interesses comuns ao aluno. Essa associação seria interessante para o professor ter em mãos ao final do curso uma lista de assuntos e palavras-chaves que cada aluno utilizou nas discussões, para identificar as ocorrências de palavras relevantes no contexto das discussões e na avaliação do aluno.

No entanto, verificamos que o uso das técnicas de *data mining* em ambientes de aprendizagem *on-line* ainda é muito restrito e que os relatos da incorporação destas técnicas nos ambientes educacionais são ainda escassos.

## Referências Bibliográficas

- Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., Zanasi, A. (1998). *Discovering Data Mining From Concept to Implementation*. Prentice Hall PTR. New Jersey, EUA.
- Fayyad, U. M. (1996). *Advances in Knowledge Discovery and Data Mining*. Menlo Park, The MIT Press.
- Nascimento, D. M. C., Rodrigues, A. H., Schiel, U. (2002). *Portfolio-Tutor: Um Tutor Acoplado a um Portfolio Eletrônico no Contexto da Educação a Distância. XIII Simpósio Brasileiro de Informática na Educação*. São Leopoldo, RS.
- Santos, N. (2003) *Agentes de Software em Ambientes Educacionais Mediados por Computador Mediados por Computador. Revista Brasileira de Informática na Educação*, Porto Alegre, Vol 2. Abril.
- Silva, D. R.; Seno, W. P.; Vieira, M. T. P. (2001). *Acompanhamento do Aprendizado em Educação a Distância com Uso de Data Mining*. In: *Conferência Latinoamericana de Informática*, Mérida, Venezuela.
- Silva, P. D. & Santos, N. (2001). *Ambientes Educacionais Mediados por Computador: Problemas e Soluções com o Uso de Agentes. Cadernos do IME Série Informática*, Rio de Janeiro, v. 12, p 11-25.
- Han, J, and Kamper, M. (2000). *Data mining: Concepts and techniques*. First Edition, New York ; Morgan Kaufmann. 2000.