Fundamentos de Mineração de Dados Educacionais

#### Leandro A. Silva, Luciano Silva

Faculdade de Computação e Informática – Universidade Presbiteriana Mackenzie Rua da Consolação, 930 - 01302-907 – São Paulo – SP – Brasil {leandroaugusto.silva,luciano.silva}@mackenzie.br

Abstract. Educational Data Mining (EDM) is a multidisciplinary field which uses Data Mining techniques in educational contexts such as virtual environments for teaching and learning. Through EDM, it is possible to infer about learners and tutors profiles, relationships among learning variables or even through analysis models. This tutorial presents the basis of MDE, starting from the fundamentals of traditional Data Mining until more complex processes involving analysis models.

**Keywords:** Educational Data Mining, Informatics in Education, Data Mining.

Resumo. Mineração de Dados Educacionais(MDE) corresponde a uma área multidisciplinar, que utiliza técnicas de Mineração de Dados em contextos educacionais como, por exemplo, ambientes virtuais de ensino e aprendizagem. Através da MDE, é possível inferir sobre perfis de aprendizes e tutores, relações entre variáveis de aprendizado ou mesmo através de modelos de análise. Este tutorial apresenta as bases de MDE, passando desde os fundamentos da Mineração de Dados tradicional até processos mais complexos envolvendo modelos de análise.

**Palavras-chave:** Mineração de Dados Educacionais, Informática em Educação, Mineração de Dados.

## 1. Introdução

Mineração de Dados (MD) é uma área de pesquisa multidisciplinar, envolvendo basicamente Banco de Dados, Estatística e Aprendizado de Máquina. A MD é parte principal de um processo que tem como entrada uma Base de Dados e como saída um Conhecimento [Fayyad et al., 1996]. Ela é divida em tarefas como predição, clusterização e associação que devem ser escolhidas de acordo com analises exploratórias inicialmente feitas sobre os dados [Han et al., 2006; Tan et al., 2009; Faceli et al., 2011]. A MD tem sido amplamente utilizada em diferentes áreas, principalmente medicina, indústria, marketing, agronegócios e muitos outras como educação [Berry & Linoff, 2004; Dunstone & Yager, 2008; Stone et al., 2008; Silva et al, 2008; Romero & Ventura, 2010].

Na Educação, ou seja, a Mineração de Dados Educacionais(MDE) é uma área de pesquisa interdisciplinar que lida com o desenvolvimento de métodos para explorar

dados oriundos de contextos educacionais [Romero & Ventura,2010; Paiva et al., 2012]. Os tipos de estudos desta área são classificados, segundo Romero & Ventura (2010) em: educação offline para analises em dados de desempenho do aluno, comportamento, currículo e etc, ou seja, gerados em ambientes de sala de aula; aprendizado eletrônico (e-learning) e Sistema de Gestão da Aprendizagem ou LMS (do inglês, Learning Management System) para analise de dados armazenados em sistemas LMS no formato de log e bases de dados; e Sistemas Tutores Inteligentes ou ITS (do inglês Intelligent Tutoring System) e Sistemas Hipermídias Adaptativos Educacionais ou AEHS (do inglês, Adaptive Educational Hypermedia System) os quais aplicados sobre dados de sistemas que se adaptam a cada estudante em particular, aos cursos oferecidos que estão em forma de log, aos modelos de usuários e etc.

Na literatura, existem diversos trabalhos que usam diferentes técnicas de Mineração de Dados no contexto educacional. Singh & Kumar (2012), por exemplo, utilizaram a técnica de mineração de dados chamada árvore de decisão para gerar conhecimento aos gestores da instituição para avaliar o desempenho de seus alunos, portanto. Dejaeger et al. (2011), por outro lado, utilizaram a técnica de mineração de dados chamada clusterização de dados para identificar os principais fatores de satisfação dos alunos em duas instituições de ensino e consequentemente para a construção de modelos para apoiar os gestores no processo de tomada de decisão estratégica.

Dentro deste contexto, o objetivo deste tutorial é apresentar as bases da MDE, iniciando com os fundamentos da Mineração de Dados tradicional até processos mais complexos como análise baseados em modelos. Este tutorial foi apresentado no início do I Workshop de Mineração de Dados Educacionais (WMDAVA), realizado em conjunto com o Simpósio Brasileiro de Informática em Educação (SBIE), em Dourados(MS).

O texto está organizado da seguinte forma: a Seção 2 apresenta os principais processos da Mineração de Dados como organização de bases de dados, préprocessamento, tarefas de Mineração de Dados e pós-processamento; a Seção 4 discute alguns exemplos de MDE envolvendo agrupamentos, classificações e modelos de análise e, finalmente, ao final do texto há um conjunto de referências bibliográficas que podem servir de apoio aos iniciantes em MDE.

Não é objetivo deste texto ser um documento denso e completo. Porém, os autores esperam que ele possa servir de base para que professores, alunos e pesquisadores interessados em MDE possam ter uma trilha mínima para iniciar pesquisas nesta área.

#### 2. Mineração de Dados

Mineração de Dados é o termo usado para uma tarefa aplicada em um grande volume de dados com o objetivo de descobrir padrões relevantes em um contexto previamente definido para apoio em tomada de decisão. Portanto, a descoberta de padrões constituise de um processo que se inicia pela escolha dos dados que documentam de alguma maneira a pergunta que o especialista deseja responder. Os dados são integrados e préprocessados para que sejam entregues estruturados, higienizados, selecionados e padronizados à tarefa de Mineração de Dados. Na tarefa de mineração aplica-se alguma

técnica inteligente capaz de encontrar soluções que auxiliam o especialista na descoberta de uma resposta. O resultado desta tarefa deve ser pós-processado para que se apresentem analises qualitativa e/ou quantitativa dos elementos encontrados e, quando possível, apresentados de maneira que possa ser interprestada de maneira a facilitar a tomada de decisão. Na Figura 1, descreve-se odo este processo, o qual pode ser chamado de Descoberta de Conhecimento em Bases de Dados ou KDD (do inglês, *Knowledge Discovery in Database*) [Fayyad et al., 1996]. Note, nesta figura de processo do KDD, que há uma dupla seta na integração de todas as etapas. Isto significa que o processo pode ser iterativo, ou seja, seguindo cada passo por vez, ou interativo, o processo pode ser executado em passos arbitrários ou mesmo repetidos em caso haja necessidades. A seguir discutiremos em mais detalhes cada uma das etapas do KDD.

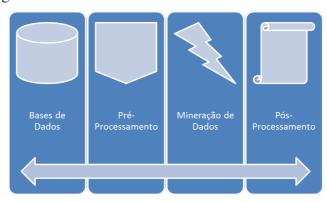


Figura 1. Processo de Descoberta de Conhecimento em Bases de Dados.

#### 2.1. Bases de Dados

O processo KDD envolve a descoberta de conhecimento de um conceito específico intrinsicamente armazenado em bases de dados. Portanto, é importante envolver neste processo um especialista que conhece o problema e que possa auxiliar na escolha de bases de dados. Isto significa que podem ocorrer casos onde será necessário utilizar mais de uma base de dados. E ainda, esta base poderá representar dados organizados em uma tabela, os quais são chamados de estruturados ou podem estar em documentos descritos na forma de texto ou multimídia (imagem, vídeo e som) os quais são chamados de não estruturados.

Nesta etapa do processo os dados devem ser integrados e assim constituir uma base de dados única. Se a pergunta que se deseja responder envolve dados não estruturados, então esses devem passar por um processo anterior e, consequentemente serem representados de maneira estruturada. No caso de mais de uma base de dados estruturada, a integração pode gerar uma serie de problemas que serão tratados na etapa seguinte de pré-processamento.

Uma base de dados é constituída por uma tabela, onde geralmente os exemplares ficam em linhas e os atributos em coluna. Não há regras que impeça a organização ser feita invertendo linhas por colunas. O exemplar é um registro da base de dados que ainda é chamado na literatura como objeto, padrão, amostra e etc. O atributo, por outro lado, descreve o exemplar. Ele ainda é chamado na literatura por características ou variáveis. Os atributos podem ser definidos em regulares e especiais. Atributos regulares são medidas feitas sobre a pergunta em analise, já os especiais é um rótulo que se dá para o exemplar. Por exemplo, em um problema de MDE onde se deseja aplicar

Mineração de Dados sobre os alunos de um curso e as notas que ele tem em todos os semestres para encontrar um modelo que seja capaz de inferir se o aluno se formará no prazo. Cada aluno é um exemplar desta base e, a nota atribuída ao aluno em cada disciplina é um atributo regular. A situação do aluno no final do curso, por exemplo, formado no prazo ou fora do prazo é um exemplo típico de atributo especial.

O valor armazenado nos atributos pode ser numérico ou categórico. O atributo numérico pode ser contínuo ou discreto. E o atributo categórico pode ser nominal ou ordinal. Estas nomenclaturas e definições são importantes para a escolha da tarefa de Mineração de Dados e os consecutivos algoritmos no problema em questão. Na seção seguinte daremos continuidade ao processo KDD, discutindo o pré-processamento.

#### 2.2. Pré-Processamento

Como discutido antes, o processo de KDD pode ser iterativo ou interativo. Portanto, a necessidade de envolver a etapa de pré-processamento deve ser avaliada a cada problema. Muitos dos casos tratados no pré-processamento são oriundos da integração de base de dados. Exemplos típicos destes casos são valores ausentes, ruidosos, inconsistentes e redundantes. Ainda há de considerar também que, em outras situações com ou sem integração de bases de dados, o aumento no número de exemplares e atributos da base de dados podem interferir na execução dos algoritmos usados nas tarefas de mineração de dados e por essa razão devem ser tratados.

Como as razões para a necessidade de pré-processamento variam muito de área de conhecimento e considerando-se que o foco deste trabalho é base de dados educacionais, contextualizaremos as causas neste domínio de interesse. Embora as causas sejam variantes ao domínio de aplicação, as soluções, por outro lado, são robustas. Isso significa que elas podem ser aplicadas não somente à MDE.

Adicionalmente às situações acima expostas, a sequência de diagnósticos de problemas nas bases de dados não deve ser entendida como ótima solução. Pode haver casos em que o tratamento pode ser feito em uma ordem contrária a aqui apresentada.

Para um diagnostico de possíveis problemas é comum fazer inicialmente uma analise descritiva dos dados. Nesta analise, medições são feitas sobre os atributos dos dados como média, desvio padrão, quartis, valor mínimo e máximo. Estas medidas auxiliam no encaminhamento da solução de pré-processamento a ser adotada e também, em caso de valores ausentes já será possível verificar a sua existência e, consequentemente a sua solução. Começaremos então a discutir este tipo de problema.

Valores ausentes ou *missing values* são atributos que não tem valores preenchidos. As razões são diversas, por exemplo, em uma integração de dados de alunos de uma faculdade envolvendo diferentes cursos, pode haver disciplinas incomuns e, portanto haverá o surgimento deste tipo de valores que precisam ser tratados. O tratamento pode ser feito pela simples remoção do atributo (em caso de grande incidência como poderia ser para o caso do exemplo apresentado) ou do exemplar (em caso de poucas ocorrências). Ou ainda o valor pode ser substituído por uma constante calculara pela média, mediana, valor máximo ou mínimo. Por fim, pode se aqui praticar a interatividade do KDD e utilizar alguma tarefa de Mineração de Dados para fazer a estimação do melhor valor para o atributo.

Outro tratamento que pode ser diagnosticado na analise descritiva são os valores ruidosos ou que estão fora do padrão (outliers). Este tipo de situação ocorre quando surge algum exemplar com valor de atributo que foge de um padrão. Isto pode ser diagnosticado com a construção de um box plot. Esse gráfico tem o valor dos 3 primeiros quartis e os valores mínimo e máximo e, portanto, pode-se verificar, visualmente a variação dos valores de cada atributo e diagnosticar a presença de outliers. Alternativamente, o desvio padrão também pode ser usado como uma informação qualitativa de valores fora de um padrão. Por exemplo, se em uma prova unificada da faculdade que historicamente tem média 5 e, esporadicamente um grupo de alunos aparecem com nota 10, isso pode significar um ruído. A razão do surgimento pode ser uma fraude (no exemplo o grupo de alunos pode ter acesso antecipado às questões da prova) ou pode ser uma mudança de comportamento no padrão da base de dados ou outlier (no exemplo, os alunos são mais dedicado que em anos anteriores). O tratamento para este tipo de ocorrência pode ser feito por diferentes abordagens [Han et al., 2006]. Tipicamente ela pode ser resolvida com a separação dos valores em faixas pré-definidas, técnica conhecida como binning [Tan et al., 2009]. Para o exemplo de contexto, o binning seria apresentar notas dos alunos não por números, mas sim por letras. Neste caso cada letra representa uma faixa de valores e assim evitam-se além do outliers efeitos de subjetividade em uma correção da prova.

Por fim, uma situação que surge tipicamente quando se faz integração de dados é a inconsistência de valores. A inconsistência ocorre quando há falta de um critério bem definido entre os valores dos atributos ou dos exemplares. Exemplo que ilustra o critério para os atributos é quando em uma analise de dados de notas de alunos e situação final, um aluno teve nota 5 e está aprovado e, na mesma base, um outro aluno tem a mesma nota 5, mas está reprovado. Isto pode ocorrer por diversas razões, como a mudança de critério de aprovação de um curso. Outra situação, agora com os valores de exemplares, considere que uma disciplina tenha aula teoria e prática. Na teórica o professor define uma escala de nota de 0 a 10 e, por outro lado, na prática outro professor de 0 a 1. Na integração destas notas haverá inconsistência entre as escalas dos valores. Para o primeiro exemplo, a solução poderia ser a remoção dos exemplares ou o ajuste dos valores, desde que se alguma informação adicional sobre os dados fornecida por um especialista. No segundo caso, a informação de valores mínimos e máximos da analise descritiva pode ajudar na solução chamada de normalização.

A normalização de valores consiste em uma técnica para deixar os valores dos atributos em uma mesma escala. No exemplo anterior, alunos com notas de 0 a 1 e de 0 até 10, a normalização faz com que se tenha uma escala única. Abordagem de solução comum é calcular o valor máximo de um atributo para dividi-lo aos demais exemplares com mesmo atributo. A normalização faz parte de um tratamento chamado transformação de valores que ainda compreende a mudança de tipos categóricos para numéricos. Ou seja, caso queira fazer uma analise mais profunda dos alunos considerando informações de escolaridade dos pais e gênero do aluno. Neste caso, a escolaridade deveria ser um atributo categórico ordinal com valores típicos segundo grau, graduação, pós-graduação que seriam transformados para 1, 2 e 3, por exemplo. No caso do gênero, atributo nominal, como são apenas dois valores, eles poderiam ser transformados para binário 0 e 1. No entanto, deve-se ter cuidado para não transformar um atributo nominal em ordinal no processo de transformação, isto é, o valor não pode ideia de ordem. Para mais de dois tipos de valores nominais, a transformação pode ser

feita com uso de sequencias de 0s e 1s, por exemplo, 01, 10, 11 e assim consequentemente.

Finalmente como última investigação feita no pré-processamento é a seleção de exemplares ou atributos. As causas que levam a se fazer este tipo de analise são muitas como, integração de bases, falta de definição clara de atributos que representam um problema, grande disponibilidade de dados e outras. A seleção consiste basicamente em escolher o melhor conjunto de dados que representam a base original com a mesma capacidade analítica. A seleção de exemplares é um processo bastante simples, basta verificar redundâncias e eliminar o exemplar que está a mais na base de dados. A seleção de atributos, por outro lado, é preciso ter um pouco mais cuidado. O que se faz basicamente é verificar o quanto dois atributos estão correlacionados e, consequentemente, escolher um deles para analise. Portanto, a correlação, por exemplo, de Pearson, é uma técnica típica para seleção de atributos, mas ainda há outras na literatura [Tan et al., 2009].

### 2.3. Tarefas de Mineração de Dados

A tarefa de Mineração de Dados é onde se extrai algum tipo de conhecimento da base de dados. Isso será feito a partir da escolha de uma tarefa e consequentemente de um algoritmo, os quais dependem basicamente dos atributos disponíveis. O conhecimento descoberto dependerá da tarefa que está em uso, que são definidas em: Modelo Preditivo, Análise de Agrupamento e Regras de Associação. Discutiremos a seguir cada uma delas.

#### **Modelo Preditivo**

A predição pode ser definida como sendo uma tentativa de se descobrir o que acontecerá em um momento futuro [Han et al., 2006]. Em Mineração de Dados, essa tentativa é feita a partir de um modelo construído usando a base de dados como referência. Por esta razão a base usada na construção do modelo é chamada de treinamento (base de treinamento) e, que para o modelo preditivo, é necessário que ela tenha o atributo especial com a classe ou valor que se deseja prever. O processo de construção do modelo recebe o nome de aprendizado e, por ter o atributo especial como referencia, designamos como sendo supervisionado. Portanto, o aprendizado supervisionado é o processo de construção de um modelo preditivo onde se conhece, para finalidades de ajustes de parâmetros dos algoritmos, o que se deseja prever [Han et al., 2006; Tan et al., 2009; Faceli et al., 2011].

Em linhas gerais, o conhecimento descoberto no modelo preditivo é a função que permite mapear um conjunto de atributos de um exemplar que não faz parte da base de treinamento em um dos valores de atributo especial. Isto significa que após o treinamento, o modelo preditivo é capaz de inferir novas situações que não estavam armazenadas na base de dados. Dependendo do tipo do atributo especial, chamamos o modelo preditivo de classificação ou previsão.

Classificação de dados é quando o atributo especial da base de treinamento é formando por uma categoria ou como também chamado na literatura por um rótulo (em inglês *label*) ou alvo (em inglês *target*). No contexto de educação, um processo de classificação poderia ser feito para prever se o aluno merece ou não uma bolsa de estudos, se ele está apto para participar de um intercambio, se ele se formará no prazo

ou não e etc. Apesar dos exemplos serem com apenas duas classes, a classificação ainda pode ser feita para múltiplas classes.

Previsão de séries é o nome dado quando o atributo especial é um valor contínuo. No contexto de educação, a previsão poderia ser feita para inferir a nota do aluno em uma disciplina, a nota média do aluno no final do curso, a quantidade de faltas no final de um semestre, dentre outros.

Na literatura encontram-se muitos algoritmos de classificação e previsão de séries. Alguns deles ainda podem ser usados para ambos os tipos de atributos especiais ou então adaptados para operarem com diferentes tipos de atributos. De maneira geral, os algoritmos são divididos basicamente em Aprendizado de Máquina e Inteligência Computacional. Exemplos típicos de algoritmos de Aprendizado de Máquina são k Vizinhos mais Próximos ou kNN (k Nearest Neighbor), Máquinas de Vetores de Suporte ou SVM (Support Vector Machine), Árvores de Decisão (Decision Tree), Bayes e muitos outros. Os algoritmos de Inteligência Computacional são aqueles inspirados na biologia, principalmente no humano [Han et al., 2006; Tan et al., 2009; Faceli et al., 2011; Witten et al., 2001]. Exemplo típico são as Redes Neurais Artificiais com algoritmos do tipo Perceptron de Múltiplas Camadas ou MLP (Multi Layer Perceptron), Funções de Base Radial (Radial Bases Function) e etc [Haykin, 2009].

### **Analise de Agrupamentos**

A descoberta de grupos ou clustering trata-se de um modelo que procura encontrar exemplares ou padrões com atributos semelhantes na base de dados. Portando, a segmentação da base em grupos é feita a partir de medidas de similaridade. Em problemas que envolvem agrupamento, não se encontra disponível o atributo especial classe ou tórulo. Por esta razão, dizemos que o aprendizado realizado pelos algoritmos de agrupamento é não-supervisionado [Jain & Dubes, 1988; Han et al., 2006; Tan et al., 2009; Jain, 2010; Faceli et al., 2011; Witten et al., 2001].

O aprendizado não-supervisionado é de grande desafio, pois não se conhece a saída que se deseja alcançar, o que significa não conhecer o número de grupos da base de dados. E ainda, os exemplares estão distribuídos em um espaço de dimensão elevada com diferentes formatos e separação. Estes aspectos demonstraram a grande dificuldade para lidar em problemas de agrupamento de dados.

Em um contexto geral, o que o agrupamento faz é a descoberta de perfil. Mais especificamente em educação, a descoberta poderia ser útil para descobrir estilos de aprendizado dos alunos, disciplinas de interesses comuns e muitas outras aplicações.

Assim como na analise preditiva, os algoritmos de agrupamento também podem ser divididos em Aprendizado de Máquina e Inteligência Computacional. Algoritmos típicos de Aprendizado de Máquina são o Agrupamento Hierárquico, k-Médias (k-Means), Agrupamento Espacial baseado em Densidade ou DBScan (*Density Based Spatial Clustering of Applications with Noise*). Por outro lado, como abordagem baseada em Redes Neurais temos o Mapa Auto-Organizável ou SOM (*Self-Organizing Map*) [Jain & Dubes, 1988; Han et al., 2006; Tan et al., 2009; Haykin, 2009; Jain, 2010; Faceli et al., 2011; Witten et al., 2001].

### Regras de Associação

A associação de dados é uma abordagem de Mineração de Dados bastante diferente das demais apresentadas. A particularidade desta tarefa é que a analise é feita a partir dos atributos regulares, os quais são chamados de itens e, a partir deles gera-se a regra de associação. Os exemplares também recebem outro nome em associação de dados, são chamados de conjunto de itens. Os algoritmos não fazem parte de nenhum tipo de aprendizado, apenas metodologias para geração de regras [Agrawal et al., 1993; Han et al., 2006; Tan et al., 2009; Jain, 2010; Faceli et al., 2011; Witten et al., 2001; Aggarwal & Reddy, 2013].

Uma regra de associação é um conhecimento que apresenta itens que ocorrem associadamente. No contexto deste trabalho, uma associação seria, por exemplo, evidenciar que alunos aprovados na disciplina A também é aprovado na disciplina B ou aprovado em A e B também é aprovado na disciplina C. Ou então, aluno que acerta uma questão X na prova também acerta a Y e assim por diante.

A associação de dados é feita em dois momentos distintos. No primeiro deles verifica-se a interseção de itens e no segundo descobre-se a regra de associação. Portanto, em casos típicos os valores dos itens são binários (numéricos ou categóricos) e o que se faz é percorrer o conjunto de itens para verificar as possíveis combinações a partir de analise de interseção. Nesta fase, valores limites são definidos para a filtragem de combinações que não ocorrem minimamente. Na literatura, os dois algoritmos mais utilizados são o A Priori e FP-Growth (*Frequent Patterns Growth*) [Agrawal et al., 1993; Faceli et al., 2011]. O resultado de ambos algoritmos é o mesmo, conjunto de itens que acontecem minimimante e em um dado conjunto. Por outro lado, na geração de regras verifica-se a probabilidade de cada item acontecer dado que um outro ocorreu. E neste caso faz-se a verificação cruzada, ou seja, no exemplo anterior, verifica-se a regra SE A ENTAO B e também, SE B ENTAO A.

#### 2.4. Pós-Processamento

Nesta última etapa do processo de descoberta de conhecimento é onde fazemos analises dos resultados. As analises podem ser quantitativa ou qualitativa, e que dependem da tarefa que foi escolhida.

O modelo preditivo é construído por um conjunto de dados rotulados. Estes dados podem ser separados em dois conjuntos, sendo geralmente chamado de treinamento e outro de teste. O conjunto de treinamento usa-se para gerar o modelo e o conjunto de teste, por outro lado, para fazer analises quantitativas. Ou seja, apresenta-se um exemplar do conjunto de teste ao modelo, verifica-se o resultado estimado e, compara-se o resultado desejado (que esta na base de dados) para quantificar desempenho. Este exemplo é uma medida típica chamada de acerto, que pode ser transformada em taxa quando multiplicamos por 100 e dividido pelo número de padrões na base de dados [Han et al., 2006; Tan et al., 2009; Faceli et al., 2011; Witten et al., 2001].

Na analise de agrupamento, por outro lado, não há valor desejado. Entretanto, mesmo assim, a quantificação pode ser feita com medidas que avaliam a coesão de um grupo e isolamento entre grupos. Mas ainda, sobretudo, a segmentação feita por um agrupamento permite também análises qualitativas. Isto deve-se ao fato que é necessário

entender o perfil dos dados presentes em cada grupo formado para se definir um conceito ou categoria [Jain & Dubes, 1988; Han et al., 2006; Tan et al., 2009; Jain, 2010; Faceli et al., 2011; Witten et al., 2001].

Em regras de associação, por sua vez, o resultado é a própria regra. No entanto, a avaliação das proposições da regra é feita no momento de sua construção. Sendo assim, elas são apresentadas já com medidas que quantificam a abrangência e confiança que se pode assumir em cada regra. Mesmo assim, ainda há analises para se fazer nestas regras, que avaliam a correlação das regras geradas [Han et al., 2006].

# 3. Exemplos de Mineração de Dados Educacionais

Nesta seção, serão apresentados alguns exemplos de aplicações de MDE encontrados tipicamente na literatura. Como referência básica para estas aplicações, recomenda-se Romero et al. (2010).

Há duas grandes categorias de técnicas em MDE:

- MDE Clássica: envolve processos de predição, agrupamento e descoberta de relações
- MDE Não-Clássica: envolve descoberta com modelos e destilação de dados para julgamento humano.

### 3.1 Predição em MDE

A predição é um processo que depende de uma janela de tempo. Existem, essencialmente, três janelas de tempo importantes em MDE:

- **Exatamente agora** ou **agora**: o estudante está usando o ambiente neste momento?
- **Futuro próximo:** um estudante que usou o ambiente há pouco tempo desenvolveu habilidades para a próxima unidade ?
- **Futuro:** qual será a nota do estudante no próximo ENEM/ENADE?

Para o processo de predição, é necessário um conjunto de atributos de entrada e, dependendo do tipo de saída desejado (categorizada ou valorada), pode-se utilizar classificadores ou regressores.

Um classificador permite uma saída categorizada e determina quais componentes e suas combinações podem predizer um determinado rótulo. A Figura 2 exibe um exemplo de classificador:

KnowledgeComp	pknow	time	totalactions	right
ENTERINGGIVEN	0.704	9	1	WRONG
ENTERINGGIVEN	0.502	10	2	RIGHT
USEDIFFNUM	0.049	6	1	WRONG
ENTERINGGIVEN	0.967	7	3	RIGHT
REMOVECOEFF	0.792	16	1	WRONG
REMOVECOEFF	0.792	13	2	RIGHT
USEDIFFNUM	0.073	5	2	RIGHT

Figura 2. Exemplo para classificador.

Na Figura 2, há um conjunto de componentes de conhecimento (KnowledgeComp) que foram adquiridas por um determinado aprendiz com certa probabilidade(pknow). A partir deste contexto, estes aprendizes foram submetidos a um conjunto de questões, que foram respondidas num determinado tempo (time), com um determinado número de ações (totalactions) e produzindo resultados corretos (RIGHT) ou incorretos (WRONG). Os resultados correspondem, exatamente, às categorias (RIGHT ou WRONG). Com o processo de predição por classificação, pode-se inferir o quanto a probabilidade de aquisição de conhecimento pode influenciar numa categoria RIGHT ou WRONG.

Já num regressor, conforme exemplo mostrado na Figura 3, ao invés de se trabalhar com categorias, utiliza-se valores numéricos:

KnowledgeComp	pknow	time	totalactions	numhints
ENTERINGGIVEN	0.704	9	1	0
ENTERINGGIVEN	0.502	10	2	0
USEDIFFNUM	0.049	6	1	3
ENTERINGGIVEN	0.967	7	3	0
REMOVECOEFF	0.792	16	1	1
REMOVECOEFF	0.792	13	2	0
USEDIFFNUM	0.073	5	2	0

Figura 3. Exemplo para regressor.

A regressão para obtenção da variável numhints foi obtida através da seguinte regressão linear:

numhints = 0.12\*pKnow + 0.932\*time + 0.11\*totalactions

Existem diversos regressores que podem ser utilizados no contexto de MDE: regressores lineares (como o exemplo), quadráticos, logarítmicos e exponenciais.

## 3.2 Agrupamento em MDE

O agrupamento em MDE, assim como em Mineração de Dados tradicional, é utilizando quando não se conhece propriedades acerca dos dados que estão sendo minerados. Por exemplo, considere-se o exemplo mostrado na Figura 4:

KnowledgeComp	pknow		time	totalactions	right
ENTERINGGIVEN	0.704		9	1	WRONG
ENTERINGGIVEN	0.502	•	10	2	RIGHT
USEDIFFNUM	0.049		6	1	WRONG
<b>ENTERINGGIVEN</b>	0.967		7	3	RIGHT
REMOVECOEFF	0.792	•	16	1	WRONG
REMOVECOEFF	0.792	ĕ	13	2	RIGHT
USEDIFFNUM	0.073	•	5	2	RIGHT

Figura 4. Exemplo para agrupamento.

No exemplo desta figura, vamos utilizar como característica comum o tempo gasto para terminar os testes: se o tempo for maior ou igual a 10, o aprendiz pertencerá ao grupo vermelho; caso contrário, pertencerá ao grupo azul.

Observe que, neste processo de agrupamento, a definição de uma característica comum foi essencial para para se construir os grupos.

### 3.3 Mineração de Relações em MDE

Na mineração de relações, o objetivo é encontrar interrelações entre os dados que estão sendo minerados. Por exemplo, considere-se o conjunto de dados mostrado na Figura 5:

KnowledgeComp	pknow		time	totalactions	right
ENTERINGGIVEN	0.704		9	1	WRONG
ENTERINGGIVEN	0.502		10	2	RIGHT
USEDIFFNUM	0.049		6	1	WRONG
ENTERINGGIVEN	0.967		7	3	RIGHT
REMOVECOEFF	0.792		16	1	WRONG
REMOVECOEFF	0.792		13	2	RIGHT
USEDIFFNUM	0.073		5	2	RIGHT

Figura 5. Exemplo para mineração de relações.

Neste exemplo, o objetivo será prospectar a probabilidade de aquisição de uma determinada componente de aprendizado (pknow) e o tempo gasto (time) para resolver questões acerca desta componente.

Neste tipo de mineração, pode-se procurar por relações lineares, quadráticas, dentre outras. É muito comum o uso de técnicas estatísticas de ajustes de dados neste contexto.

#### 3.3 MDE com Base em Modelos

Num processo mais complexo, a MDE pode ser baseada em modelos para se realizar uma análise direcionada dos dados. Um exemplo típico desta técnica é o Rastreamento Bayesiano do Conhecimento, utilizando o modelo de aprendizagem de Corbett e Anderson.

O esquema geral deste modelo é mostrado na Figura 6:

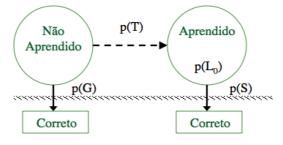


Figura 6. Esquema geral do Modelo de Corbett e Anderson.

Neste modelo, um determinado aprendiz pode estar em dois contexto: não aprendeu um determinado assunto ou aprendeu este assunto. Há uma probabilidade p(T)

......

de haver uma transição entre o estado de Não-Aprendido para o estado de Aprendido. Em ambos os estados, trabalha-se com duas probabilidades:

- p(G): probabilidade do aprendiz responder corretamente um grupo de questões dado que não aprendeu o conteúdo (por exemplo, o aluno "chutou" a resposta);
- p(S): probabilidade do aprendiz responder corretamente um grupo de questões dado que aprendeu o conteúdo.

Adicionalmente, há uma quarta probabilidade,  $p(L_0)$ , que especifica a probabilidade de um aprendiz possuir um conhecimento (um conhecimento prévio, por exemplo) no estado de aprendizado de um determinado conteúdo.

Com base neste modelo, pode-se rastrear o conhecimento numa cadeia de respostas  $corretas(Correct_n)/incorretas(Incorrect_n)$  em um determinado instante n, em função de conhecimentos anteriores  $(L_{n-1})$  utilizando as fórmulas abaixo:

$$P(L_{n-1}|\textit{Correct}_n) = \frac{P(L_{n-1}) * (1 - P(S))}{P(L_{n-1}) * (1 - P(S)) + (1 - P(L_{n-1})) * (P(G))}$$

$$P(L_{n-1}|Incorrect_n) = \frac{P(L_{n-1}) * P(S)}{P(L_{n-1}) * P(S) + (1 - P(L_{n-1})) * (1 - P(G))}$$

A primeira fórmula especifica qual a probabilidade do aprendiz ter adquirido um conhecimento anterior (n-1) dado que, no passo n, ele produziu uma resposta correta para uma determinada questão. De maneira análoga, interpreta-se a segunda fórmula.

As probabilidades  $P(L_{\scriptscriptstyle n-1})$  presentes nas fórmulas poderiam ser as colunas pknow dos exemplos anteriores.

### 3.4 Destilação de Dados para Julgamento Humano

Seres humanos podem realizar inferências além do escopo possível para sistemas automatizados de mineração de dados como, por exemplo, o raciocínio com contradição da Lógica Paraconsistente. Assim, a destilação de dados resume e apresenta os dados de maneira útil, interativa e com apelo visual para favorecer a identificação e classificação de padrões por seres humanos. Um exemplo típico desta técnica são as bem-conhecidas curvas de aprendizagem, conforme mostrado na Figura 7:



Figura 7. Exemplo de Curva de Aprendizagem.

A simples análise desta figura permite que um ser humano conclua que, em média, o aprendizado aumenta com a experiência.

......

#### 4. Comentários Finais

O objetivo primordial deste texto, assim como a própria realização do I Workshop de Mineração de Dados em Ambientes Virtuais de Aprendizagem (WMDAVA) dentro do contexto do Simpósio Brasileiro de Informática em Educação, foi promover um ambiente para apresentação e discussão de trabalhos em MDE no Brasil. Os autores esperam que os assuntos tratados aqui, bem como no próprio workshop, possam iniciar uma longa colaboração entre alunos, professores e pesquisadores.

#### Referências

- Aggarwal, C. C., & Reddy, C. K. (2013). Data Clustering: Algorithms and Applications. USA: CRC Press.
- Agrawal, R., Imieliński, T., & Swami, A. (1993, June). Mining association rules between sets of items in large databases. In ACM SIGMOD Record(Vol. 22, No. 2, pp. 207-216). ACM.
- Berry, M. J., & Linoff, G. S. (2004). Data mining techniques: for marketing, sales, and customer relationship management. John Wiley & Sons.
- Dejaeger, K., Goethals, F., Giangreco, A., Mola, L. and Baesens, B. (2011), "Gaining insight into student satisfaction using comprehensible data mining techniques", European Journal of Operational Research, Vol. 218(2), pages. 548-562.
- Dunstone, T., & Yager, N. (2008). Biometric system and data analysis: Design, evaluation, and data mining. springer.
- Faceli, K., Lorena, A. C., Gama, J., & Carvalho, A. C. P. L. F. (2011). Inteligência Artificial: Uma Abordagem de Aprendizagem de Máquina, LCT.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM, 39(11), 27-34.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM, 39(11), 27-34.
- Han, J., Kamber, M., & Pei, J. (2006). Data mining: concepts and techniques. Morgan kaufmann.

- Haykin, S. (2009). Neural networks and learning machines (Vol. 3). Upper Saddle River: Pearson Education.
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8), 651-666.
- Jain, A. K., & Dubes, R. C. (1988). Algorithms for clustering data. New Jersey: Prentice-Hall.
- Paiva, R., Bittencourt I. I., Pacheco H., Silva A. P., Jacques P., Isotani S. "Mineração de dados e a gestão inteligente da aprendizagem: desafios e direcionamentos", XXXII Congresso da Sociedade Brasileira de Computação, 2012, Curitiba. Anais do DEsafIE! I Workshop de Desafios da Computação Aplicada à Educação, 2012.
- Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 40(6), 601-618.
- Silva, L. A., Rangayyan, R. M., & Hernandez, E. D. M. (2008). Classification of breast masses using a committee machine of artificial neural networks. Journal of Electronic Imaging, 17(1), 013017-013017.
- Singh, S. and Kumar, V. (2012), "Classification of Student's data Using Data Mining Techniques for Training & Placement Department in Technical Education", (IJCSN) International Journal of Computer Science and Network, Vol. 1(4), pages.121-126, ISSN: 2277-5420, 2012.
- Stone, Z., Zickler, T., & Darrell, T. (2008, June). Autotagging facebook: Social network context improves photo annotation. In Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on (pp. 1-8). IEEE.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2009). Introdução ao datamining: mineração de dados. Rio de Janeiro: Ciência Moderna.
- Witten, I. H., Frank, E., & Mark, A. (2011). Hall (2011)." Data Mining: Practical machine learning tools and techniques.