

Data Mining – Mineração de Dados Uma Abordagem Geral

Michelly Martins Ferreira, Henrique Nunes Pereira Oliva, Janio Rodrigues Lopes, Mariane Luyara Campos Magalhães, Mikael Patrick Santos de Melo, Ted Ronaldo Aguiar Ramos, Déborah Lorranny Rodrigues dos Santos

INTRODUÇÃO

O armazenamento de dados em sistemas computacionais tem se tornado uma prática comum e essencial hoje em dia, um fator que conspira a favor disto é a queda no custo do armazenamento de dados e a rápida informatização das empresas, que facilita para o crescimento de dados armazenados, seja em data warehouses, banco de dados e outros tipos de armazenamento, de maneira centralizada ou distribuída. (REZENDE, 2005).

O conceito de Data Mining está se tornando cada vez mais popular como uma ferramenta de gerenciamento de informação, que deve revelar estruturas de conhecimento, que possam guiar decisões em condições de certeza limitada. Recentemente, tem havido um interesse crescente em desenvolver novas técnicas analíticas, especialmente projetadas para tratar questões relativas à Data Mining. No entanto, Data Mining ainda está baseada em princípios conceituais de Análise de Dados Exploratórios e de modelagem (PUC, 2013).

A Mineração de Dados

A mineração de dados pode ser considerada como uma parte do processo de Descoberta de Conhecimento em Banco de Dados (KDD – Knowledge Discovery in Databases). O termo KDD é usado para representar o processo de tornar dados de baixo nível em conhecimento de alto nível, enquanto mineração de dados pode ser definida como a extração de padrões ou modelos de dados observados (ALMO, 2004).

O KDD é um processo mais amplo onde constitui as etapas: limpeza dos dados; integração dos dados; seleção; transformação; mineração; avaliação e visualização (ALMO, 2004). Como ilustrado na Figura 1.

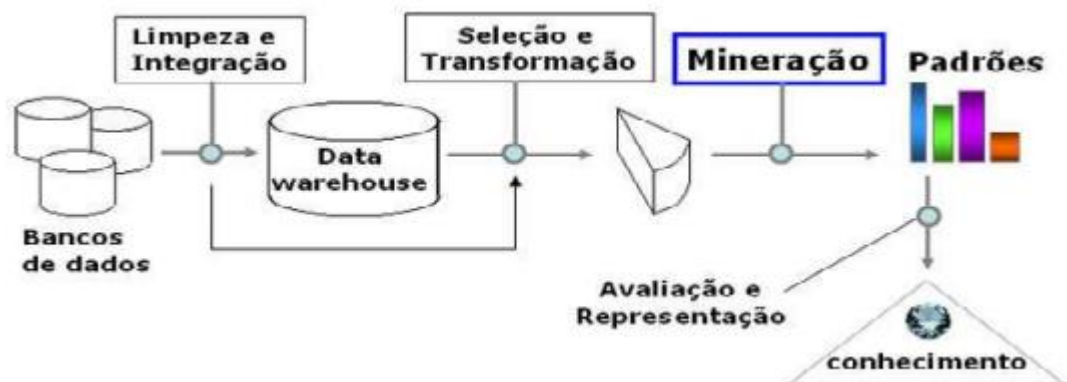


Figura 1 - Etapas do processo KDD
FONTE: ALMO, 2014

Principais tarefas

Em geral, as tarefas do *data mining* podem ser classificadas em duas categorias: *descriptive data mining* e *predictive data mining*. O primeiro descreve o conjunto de dados de uma maneira concisa e resumida e apresenta propriedades gerais interessantes dos dados; o segundo constrói um conjunto de modelos, realiza inferências sobre o conjunto de dados disponíveis e tenta prever o comportamento de novos conjuntos de dados.



FÓRUM ENSINO • PESQUISA
EXTENSÃO • GESTÃO

FEPEG

UNIVERSIDADE: SABERES E PRÁTICAS INOVADORAS

Trabalhos científicos • Apresentações artísticas e culturais • Debates • Minicursos e Palestras

REALIZAÇÃO:



Unimontes
Universidade Estadual de Montes Claros

APOIO:



FAPEMIG



FADENOR

24 a 27
setembro

Campus Universitário Professor Darcy Ribeiro

www.fepeg.unimontes.br

Um sistema de *data mining* pode realizar pelo menos uma das seguintes tarefas: a) Descrição de classes; b) Associação; c) Classificação; d) Previsão; e) Agrupamento e f) Análise de série temporal (LUVIZÃO,2006).

Técnicas de mineração de dados mais adequadas

A escolha das técnicas de mineração de dados dependerá da tarefa específica a ser executada e dos dados disponíveis para análise. A seleção das técnicas de mineração de dados deve ser dividida em dois passos: 1) traduzir o problema de negócio a ser resolvido em séries de tarefas de mineração de dados e 2) compreender a natureza dos dados disponíveis em termos de conteúdo e tipos de campos de dados e estrutura das relações entre os registros (HARRISON, 1998). Essa escolha pode ser baseada, também, em critérios para classificação das técnicas (HARRISON, 1998).

Diferentes esquemas de classificação podem ser usados para categorizar métodos de mineração de dados sobre os tipos de bancos de dados a serem estudados, os tipos de conhecimento a serem descobertos e o tipo de técnicas a serem utilizadas como segue segundo CHEN (1996):

- a) Com que tipos de bancos de dados trabalhar: Um sistema de descoberta de conhecimento pode ser classificado de acordo com os tipos de bancos de dados sobre os quais técnicas de mineração de dados são aplicadas, tais como: bancos de dados relacionais, bancos de dados de transação, orientados a objetos, dedutivos, espaciais, temporais, de multimídia, heterogêneos, ativos, de herança, banco de informação de Internet e bases textuais;
- b) Qual o tipo de conhecimento a ser explorado: Vários tipos de conhecimento podem ser descobertos por extração de dados, incluindo regras de associação, regras características, regras de classificação, regras discriminantes, agrupamento, evolução e análise de desvio;
- c) Qual tipo de técnica a ser utilizada: A extração de dados pode ser categorizada de acordo com as técnicas de mineração de dados subordinadas. Por exemplo, extração dirigida a dados, extração dirigida a questionamento e extração de dados interativa. Pode ser categorizada, também, de acordo com a abordagem de mineração de dados subordinada, tal como: extração de dados baseada em generalização, baseada em padrões, baseada em teorias estatísticas ou matemáticas, abordagens integradas e etc.

Áreas de aplicação de técnicas de mineração de dados

As áreas de aplicação de técnicas de mineração de dados podem ser as mais diversas possíveis, a seguir, são citadas as principais áreas de interesse na utilização de mineração de dados: a) Marketing; b) Detecção de fraudes; c) Medicina; d) Instituições governamentais; f) Controle de processos e controle de qualidade; g) Banco; h) Apólice de seguro; i) Transporte; h) C & T (Ciência e Tecnologia); j) Web (CRATOCHVIL, 1999).

Limitações

Apesar da grande potencialidade oferecida pela Mineração de Dados, alguns fatores devem ser analisados. É posto em discussão como alguns desses fatores podem prejudicar as técnicas de mineração: a) As relações entre os atributos precisam ser muito bem definidas, caso contrário os resultados podem ser mal interpretados; b) Permitir que o processo de treinamento execute por muito tempo, até que se consiga obter indícios que possam levar à conclusões factíveis; c) Gerar subsídios para uma conclusão errada tornando-a mais plausível. Porém, uma interpretação falha pode disfarçar as falhas nos dados; d) Usar um grande número de variáveis (WANG, 2006).

CONSIDERAÇÕES FINAIS

Devido o crescimento da tecnologia de armazenamento e a quantidade excessiva de dados que as empresas estão produzindo, o armazenamento e a análise manual destes dados se tornou inviável. Com isto a Mineração de Dados tem sido uma ferramenta de apoio com papel fundamental na gestão da informação

dentro das organizações. As empresas estão utilizando a mineração de dados para obter conhecimento sobre os seus produtos, clientes, parceiros e etc., isto para adquirir vantagem competitiva no mercado.

As técnicas de mineração passaram a estar presentes no dia a dia. Os dados são considerados hoje como o principal ativo de um projeto de software. Isso se deve, além da redução nos custos de aquisição de hardware e software, ao desenvolvimento de técnicas capazes de extrair, de forma otimizada, a informação contida, e muitas vezes implícita, nestes dados.

Bons resultados foram obtidos com a aplicação da Mineração de Dados, mas os desafios ainda são muitos. Problemas ligados ao uso da mineração, como a segurança dos dados, a privacidade dos indivíduos, juntamente com o aumento na complexidade das estruturas de armazenamento, criam cenários complexos e desafiadores.

A mineração de dados está a cada dia se popularizando. As ferramentas estão mais amigáveis e fáceis de serem usadas por usuários que não sejam especialistas em mineração, isto desempenha um papel fundamental nesse sentido. O conhecimento de todos sobre esta ferramenta é fundamental para o crescimento e a consolidação da Mineração de Dados.

A área da mineração de dados se mostra extremamente promissora, e apesar dos resultados já obtidos, ainda tem muito para oferecer.

REFERÊNCIAS

REZENDE, SOLANGE OLIVEIRA. **Mineração de Dados**. 2005. Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/enia/2005/0102.pdf>>. Acesso em: 20 Jul. 2014.

PUC. **Data Mining – Conceitos, Técnicas, Ferramentas e Aplicações**. 2013. Disponível em: <<http://www.cce.puc-rio.br/sitecce/website/website.dll/folder?cOferec=7003>>. Acesso em: 02 Jul. 2014.

AMO, S. **Técnicas de Mineração de Dados**. 2004. Universidade Federal de Uberlândia. Disponível em: <<http://www.lsi.ufu.br/documentos/publicacoes/ano/2004/JAI-cap5.pdf>>. Acesso em 15 Jul. 2014.

LUVIZÃO, L. *et al.* **Aspectos Teóricos do Data Mining e Aplicação das Redes Neurais em Previsões de Preços Agropecuários**. 2006. Universidade Federal de Viçosa. Disponível em: <<http://ageconsearch.umn.edu/bitstream/148202/2/681.pdf>>. Acesso em 01 Jul.2014.

HARRISON, Thomas H. **Intranet Data Waregouse**. São Paulo, Berkeley Brasil, 1998.

CHEN, M.S. *et al.* **Data mining: an overview from database perspective**. IEEE Transactions on Knowledge and Data Engineering, v. 8, n.6, p. 866-883, 1996.

CRATOCHVIL, A. **Data mining techniques in supporting decision making**. Master Thesis - Universiteit Leiden, Leiden, 1999.

WANG, Q. *et al.* **Strategic postures of third-party logistics providers in mainland China**. *International Journal of Physical Distribution & Logistics Management*, v. 36, n. 10, p. 793 - 819, 2006.