
2006. 2

Administração de dados - Conceitos, técnicas,
ferramentas e aplicações de Data Mining para gerar
conhecimento a partir de bases de dados

PROPOSTA DE TRABALHO DE
GRADUAÇÃO EM
BANCO DE DADOS

Aluno: Thiago Amorim (tmas@cin.ufpe.br)

Orientador: Fernando da Fonseca de Souza (fdfd@cin.ufpe.br)

Contexto

Cada vez mais, imensos volumes de informação têm sido sistematicamente coletados e armazenados por grandes organizações. E o custo de uma má qualidade de dados, pode ser decisivo para o sucesso de uma empresa. Com isso, é essencial o estudo da Administração de Dados para garantir a qualidade dos dados que são inseridos e recuperados de uma base de dados. Administração de dados é o desenvolvimento e execução de estratégias, práticas e procedimentos para o processo de gerência do ciclo de vida completo dos dados em uma empresa.

As organizações têm se mostrado extremamente eficientes em capturar, organizar e armazenar grandes quantidades de dados, obtidos de suas operações diárias ou pesquisas científicas, porém, a maioria delas ainda não usa adequadamente essa gigantesca montanha de dados para transformá-la em conhecimentos que possam ser utilizados em suas próprias atividades. Com a geração de um volume cada vez maior de informação, é essencial tentar aproveitar o máximo possível desse investimento.

Uma excelente prática de Administração de Dados é o enriquecimento dos dados, gerando ainda mais informação e conhecimento, melhorando assim a sua qualidade. Data Mining é uma etapa de um processo de enriquecimento de dados conhecido como extração de conhecimento em bases de dados ou Knowledge-Discovery in Databases (KDD). Uma definição importante para KDD foi elaborada por Usama Fayyad (Fayyad et al. 1996): "...o processo não-trivial de identificar, em dados, padrões válidos, novos, potencialmente úteis e ultimamente compreensíveis". [1]

Mineração de dados, ou data mining, é o processo de análise de conjuntos de dados que tem por objetivo a descoberta de padrões interessantes e que possam representar informações úteis. O processo de Data Mining permite que se investigue esses dados à procura de padrões que tenham valor para a empresa. 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.

Com a utilização de data mining, pode-se analisar dados comportamentais, obtendo conhecimento que estava "escondido" na base de dados, gerando aprendizado e dados complementares que podem influenciar no desenvolvimento de estratégias na organização.

Objetivo

Neste contexto, este trabalho de graduação se propõe a especificar e detalhar cada fase do processo de data mining para geração de conhecimento a partir de bases de dados. Desde a fase de obtenção dos dados brutos e sua "simplificação" ou "limpeza", abstraindo-os em dados mais genéricos; passando pela fase de reconhecimento dos padrões, utilizando-se várias técnicas e algoritmos; até a fase de avaliação da validade e da importância dos conhecimentos gerados, avaliando o quanto bom ou interessante é cada padrão, para decidir o que deve ser mantido, o que deve ser descartado e o que deve ser melhor explorado.

Também serão demonstradas várias situações práticas da utilização de data mining em variados tipos de organizações, mostrando que a Administração de Dados não só garante uma segurança e confiabilidade dos dados, como também pode trazer grandes benefícios extras para uma empresa.

Por fim, será apresentada uma análise de algumas ferramentas de data mining disponíveis no mercado, descrevendo suas funcionalidades assim como as principais características de cada uma.

Estrutura do Trabalho

O trabalho de graduação será estruturado da seguinte maneira:

O **capítulo 1** será uma introdução. Nesta seção será estabelecido o contexto em que se aplica o trabalho, como também a motivação para o seu desenvolvimento. Serão citados os objetivos a serem alcançados e será apresentada a estrutura completa do trabalho.

O **capítulo 2** irá detalhar todo o processo de mineração de dados, detalhando cada fase desse processo, desde a seleção dos dados e sua filtragem, aos algoritmos e técnicas utilizadas para o reconhecimento de padrões, como também os meios de medir a validade e importância do conhecimento gerado.

O **capítulo 3** mostrará o uso do processo de mineração de dados em diferentes áreas de conhecimento, mostrando o quanto essa técnica pode agregar de valor para uma empresa, e para a sociedade.

O **capítulo 4** apresentará um estudo comparativo de ferramentas de data mining disponíveis atualmente, mostrando as funcionalidades e características de cada uma delas, como também as técnicas que elas utilizam.

O **capítulo 5** irá apresentar uma conclusão para o trabalho, mostrando as contribuições providas pelo mesmo, como também apresentando sugestões de possíveis trabalhos futuros.

O **capítulo 6** irá listar a referência bibliográfica, para um estudo mais aprofundado das fontes de pesquisa deste trabalho.

Por fim, se necessário, haverá a inclusão de apêndices em anexo, com algumas informações extras, ou detalhamento de partes citadas no escopo do trabalho.

Referência

[1] Knowledge Discovery in Databases: Tools and Techniques

<http://www.acm.org/crossroads/xrds5-2/kdd.html>

Bibliografia

Princípios Essenciais do Data Mining:

<http://www.intellicore.com/reports/i2002.htm>

Data Mining Tutorials, Resources:

<http://datamining.eruditiohome.com/>

Data Mining: um Guia Prático

Ronaldo Goltschmidt, Emmanuel Passos

Data Mining – Michel de Souza

http://www.imasters.com.br/artigo/1482/bi/data_mining/

Advances In Knowledge Discovery And Data Mining

U. M. Fayyad, G. Piattetsky-Shapiro, P. Smyth e R. Uthurusamy

Data e Assinaturas

Thiago Miranda Amorim Silva
(Aluno)

Fernando da Fonseca de Souza
(Orientador)

Recife, 28 de Novembro de 2006