

USP

LABIC

# Data Mining

## Uma Visão Geral

**Solange Oliveira Rezende**

USP São Carlos / ICMC  
 Departamento de Ciências de Computação e Estatística  
 Laboratório de Inteligência Computacional  
<http://www.icmc.usp.br>

© Solange Oliveira Rezende

USP

LABIC

# Tópicos

## Curso de Aprendizado de Máquina

Abril 2001

Profa. Dra. Maria Carolina Monard

**LABIC - ICMC - USP São Carlos**

USP

LABIC

# Tópicos do curso de AM

- ♦ Inteligência e Aprendizado
- ♦ Engenharia de Conhecimento e SBC
- ♦ Dado, Informação e Conhecimento
- ♦ Aprendizado de Máquina
- ♦ AM usando Árvores de Decisão
- ♦ KDD e Data Mining

USP

LABIC

# Sistemas de IA

```

      graph TD
        A([Conceitos e técnicas de IA]) --> B[Sistemas de IA]
        B <--> C[Conhecimento]
        C --- D((Processo de Aprendizado))
        D --- E[Aquisição]
        D --- F[Representação]
        D --- G[Manipulação]
      
```

USP

LABIC

# Sistemas de IA

```

      graph TD
        A([Conceitos e técnicas de IA]) --> B[Sistemas de IA]
        B <--> C[Conhecimento]
        C --- D((Processo de Aprendizado))
        D --- E[Aquisição]
        D --- F[Representação]
        D --- G[Manipulação]
        H([Data Mining]) --- I[Aprendizado de Máquina]
        I --- D
      
```

USP

LABIC


# Aprendizado de Máquina

Pode ser utilizado como meio para vencer um dos maiores problemas de Sistemas de IA - o gargalo da aquisição de conhecimento

USP

LABIO

## Aprendizado de Máquina

 Sub-área da Inteligência Artificial que pesquisa métodos computacionais relacionados à aquisição de novos conhecimentos, novas habilidades e novas formas de organizar o conhecimento já existente.

SOF

USP

LABIO

## Objetivos de AM

- ♦ Um melhor entendimento dos mecanismos de aprendizado humano
- ♦ Automação da aquisição do conhecimento

SOF

USP

LABIO

## AM incorpora várias técnicas de outras disciplinas



SOF

USP

LABIO

## Sistemas de Aprendizado de Máquina


Modo de Aprendizado	Paradigmas de Aprendizado	Linguagens de Descrição	Formas de Aprendizado
- Supervisionado	- Simbólico	- Instâncias ou Exemplos	- Incremental
- Não Supervisionado	- Estatístico	- Conceitos	- Não Incremental
	- Instance-Based	- Aprendidos ou Hipóteses	
	- Conexionista	- Teoria de Domínio ou Conhecimento de Fundo	
	- Genético		

SOF

USP

LABIO

## Data Mining x Aprendizado de Máquina



Muitos Exemplos    Poucos Exemplos    Poucos Atributos    Muitos Atributos


AM geralmente (mas não sempre) trabalha com pequena quantidade de dados, mas relevantes

SOF

USP

LABIO

## Data Mining x Aprendizado de Máquina



Muitos Exemplos    Muitos Atributos

DM trabalha com grandes bases de dados reais, sendo a eficiência muito importante

SOF

USP

LABIO

## Tópicos

- ♦ Revisão do curso de AM (Profa. Carolina)
- ♦ Introdução
- ♦ Dado, Informação, Conhecimento, ...
- ♦ Data Mining
- ♦ Etapas do Processo de Data Mining
- ♦ Tecnologias que dão suporte para DM
- ♦ Técnicas e Ferramentas para DM
- ♦ Tendências
- ♦ Algumas Aplicações
- ♦ Estudos de Caso
- ♦ Considerações Finais

SOF

USP

LABIO

## Conhecimento para Apoio à Tomada de Decisão

SOF

USP

LABIO

## Introdução

Poder de processamento (aumento)  
Armazenamento de dados (baixo custo)  
Novas tecnologias de transmissão de dados  
Popularização da Internet

SOF

USP

LABIO

## De Dados a Manipulação de Conhecimento: uma Estrutura

SOF

USP

LABIO

## O que é DADO?

Fatos, imagens ou sons que podem ou não ser úteis para uma determinada tarefa.

- ♦ Dado é a estrutura fundamental sobre a qual um sistema de informação é construído.
- ♦ Ex: AKWHGGAFWPPH

SOF

USP

LABIO

## O que é a INFORMAÇÃO?


Dados cuja forma e conteúdo são úteis para uso no processo de tomada de decisão.

- ♦ Parte do processo de criar informação é generalizado a partir de base de dados.
- ♦ Ex: AKWHGGAFWPPH – representa parte de um sequenciamento

SOF

USP Dado, Informação, Conhecimento LABIO

## O que é CONHECIMENTO?

 Fornece a capacidade de resolver problemas, inovar e aprender baseado em experiências prévias.

- ♦ Uma combinação de instintos, idéias, regras e procedimentos que guiam as ações e decisões.
- ♦ O êxito do conhecimento em explicar a forma como as coisas acontecem é um critério importante na validação desse conhecimento.

SOF

USP Dado, Informação, Conhecimento LABIO

## Importante observar ....

Dado não é Informação  
 Informação não é Conhecimento  
 Conhecimento não é Inteligência  
 Inteligência não é Sabedoria

SOF

USP Ex: Dado - Informação - Conhecimento LABIO

- ♦ 300 - mm de chuva - >200mm -> inundación
- ♦ 38 - pontos de independente - campeon
- ♦ 10:15 - hora - coffe break


SOF

USP Evolução da Tecnologia LABIO

EVOLUÇÃO	TECNOLOGIA UTILIZADA
Coleta de dados (1960s)	Computadores, Fitas, Discos
Acesso aos Dados (1980s)	RDBMS, SQL, ODBC
Navegação pelos dados (1990s)	OLAP, Base de Dados Multidimensionais, Data Warehouse
Data Mining (2000)	Algoritmos Avançados, Computadores com Multiprocessadores, Grandes Bases de Dados

SOF

USP Definição LABIO



Processo de identificação de padrões válidos, inovadores, potencialmente úteis e, principalmente, compreensíveis em conjuntos de dados (Fayyad et al., 1996)

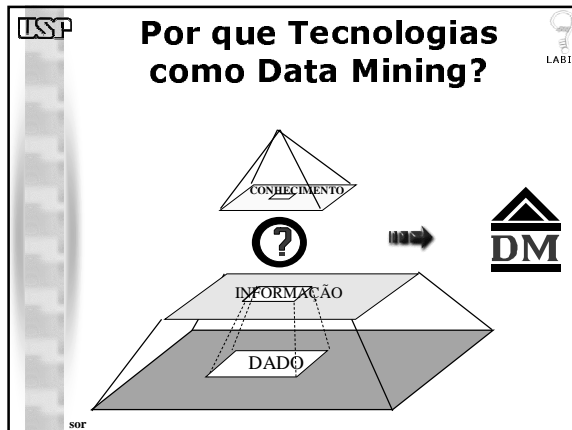
SOF

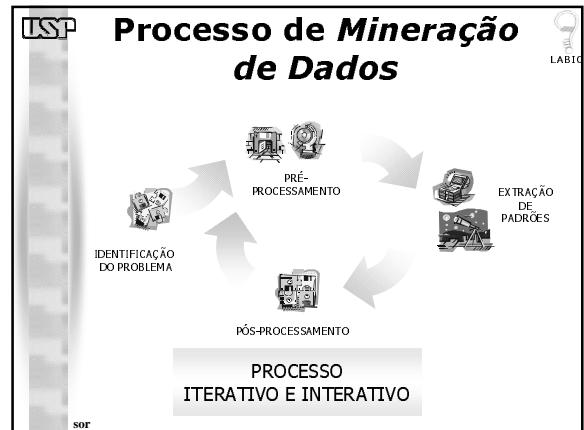
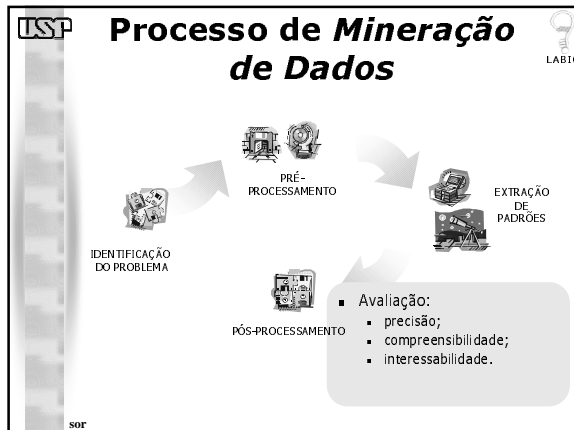
USP Alguns Casos de Sucesso LABIO

NIKE  
  → 

WAL MART  
  → 

SOF





- USP** **Pontos importantes em DM...** **LABIO**
- ♦ Dado cru é fácil de coletar, mas caro para analisar.
  - ♦ Métodos: aprendizado de máquina, estatística, bases de dados + visualização.
  - ♦ Uma ferramenta para análises de dados.
  - ♦ Pode ser combinado com métodos tradicionais.
  - ♦ Grande interesse desde 1989.
  - ♦ Sucesso nas aplicações práticas.
- SOF**

- USP** **Algumas Considerações...** **LABIO**
- ♦ Usualmente percebe-se uma grande confusão de termos na comunidade como: Data Warehouse, OLAP, Data Mining, KDD
  - ♦ KDD (*Knowledge Discovery in Databases*) e DM (Data Mining) são muitas vezes utilizados como sinônimo.
  - ♦ É comum utilizar DM em contextos industriais, enquanto que KDD é uma expressão mais científica.
- SOF**

- USP** **Algumas Considerações... (cont)** **LABIO**
- ♦ **A indústria vê Data Mining como:**
    - ✧ Apoio a Decisão
    - ✧ Bussiness Inteligence
    - ✧ Próximo passo após o Data Warehouse...
- SOF**

## Alguns Livros Básicos

- ♦ Rezende, S. O., J. B. Pugliesi, E. A. Melanda, & M. F. Paula (2003). Mineração de dados. In S. O. Rezende (Ed.), *Sistemas Inteligentes: Fundamentos e Aplicações*, pp. 311 – 339. Ed. Manole.
- ♦ Fayyad, U. M. ; Piatetsky-Shapiro, G.; Smyth, P.; Uthurusamy, R. *Advances in Knowledge Discovery and Data Mining*, MIT Press, 1996.
- ♦ Witten, I. H.; Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann, 1999.  
<http://www.cs.waikato.ac.nz/~ml/weka/book.html>
- ♦ Pyle, D. *Data Preparation for Data Mining*, Morgan Kaufmann Publishers, 1999.
- ♦ Thuraishingham, B. *Data Mining: Technologies, Techniques, and Trends*, CLR Press LLC, 1999.

## Próximos Tópicos

- ♦ Etapas do Processo de Data Mining
- ♦ Tecnologias que dão suporte para DM
- ♦ Técnicas e Ferramentas para DM
- ♦ Tendências
- ♦ Algumas Aplicações
- ♦ Estudos de Caso
- ♦ Considerações Finais