

Inteligência Artificial

Aula 24- Aprendizagem de Máquina: Introdução ¹

Sílvia M.W. Moraes

Faculdade de Informática - PUCRS

May 30, 2017

¹Este material não pode ser reproduzido ou utilizado de forma parcial sem a permissão dos autores.

Sinopse

- Nesta aula, introduzimos em **aprendizagem de máquina**.
- Este material foi construído com base no material sobre Data Mining dos professores Rodrigo Barros, Duncan e Renata de Paris e também nos capítulos:
 - 1,2 e 3 - Inteligência Artificial: Uma abordagem de Aprendizagem de Máquina: Facelli e outros.
 - 10 do livro Inteligência Artificial: Luger
 - 18 do livro Artificial Intelligence – a Modern Approach: Russel & Norvig

Sumário

- 1 O que vimos ...
- 2 Introdução
- 3 Paradigmas de Aprendizagem
- 4 Descoberta de Conhecimento em Bases de Dados

Aulas anteriores

- Agente Reativos e Cognitivos
- Solução de Problemas: Algoritmos de busca
- Planejamento Clássico
 - Linguagens: STRIPS e PDDL
 - Planejamento Total x Parcial
 - Grafo de planejamento
- Introdução à Raciocínio Probabilístico
 - Redes Bayesianas

Motivação

- A capacidade de **aprendizagem** é essencial para um **comportamento inteligente**.
- Feigenbaum e McCorduck, em 1993, identificaram o “**gargalo da engenharia do conhecimento**”, como o maior obstáculo para o uso de larga escala de sistemas inteligentes.
 - **Revisando:** Dado x Informação x Conhecimento

Revisando: O que é dado ?

- A noção de dado está associada aos valores não interpretados que o sistema possui para processamento (Solange, 2005).
 - **Dado:** elemento puro, quantificável sobre um determinado evento que por si só não oferece embasamento para o entendimento de uma situação.
 - **Exemplos:**
 - cotação do dólar,
 - número de alunos,
 - média de idade.

Revisando: O que é informação ?

- A noção de informação está relacionada a dados selecionados e organizados para um fim determinado.
 - **Informação:** “dado analisado e contextualizado”, envolve interpretação de um conjunto de dados em um contexto de referência (Solange, 2005).
 - Exemplos:
 - Em 2012, a cotação do dolar teve uma queda histórica.
 - O número de alunos matriculados na disciplina de IA do SI aumentou em 2011.
 - A média de idade caiu entre os mestres nos últimos 5 anos.

Revisando: O que é conhecimento ?

- Conhecimento: é mais que informação.
 - **Conhecimento** é informação estruturada. A estruturação permite construir procedimentos para explorar as relações entre dados, atribuindo-lhes semântica (Solange, 2005).
 - Exemplos:
 - O dolar baixou consideravelmente devido a uma crise americana no mercado imobiliário.
 - O número de alunos matriculados aumentou porque em 2011 houve uma mudança curricular no curso de SI. Essa mudança eliminou vários pré-requisitos. A ausência de pré-requisitos permitiu que um número maior de alunos a cursasse.
 - A média de idade caiu nos últimos 5 anos, em função de vários incentivos federais em pós-graduação para recém formados, bem como devido a novas exigências do MEC.

Conceito de Aprendizagem

O que significa aprender ?



Conceito de Aprendizagem

O que significa aprender ?



*“A **mera observação** de algo **não tem valor** algum.*

***Observar** se transforma **em notar**,*

***notar** se transforma **em pensar**,*

***pensar** se transforma **em estabelecer conexões**,*

*de forma que podemos dizer que **todo o olhar atento** que lançamos sobre o mundo **é um ato de teorizar**.*

*Contudo, isso **deve ser feito conscientemente**, com **autocrítica**, com **liberdade** e, para usar uma palavra audaciosa, com **ironia**”*

- Goethe

Conceito de Aprendizagem

O que significa aprender ?

- **aprender**

- adquirir conhecimento (de), a partir de estudo; instruir-se.
Ex: "aprender uma língua, uma técnica, uma ciência"
- adquirir habilidade prática (em).
Ex: "aprendeu um esporte"

- **aprendizado**

- ato, processo ou efeito de aprender; aprendizagem.
- experiência inicial do que se aprendeu; prática, experiência, aprendizagem.

Conceito de Aprendizagem

- **Aprender envolve atividades como :**
 - **memorizar, observar e explorar** situações para aprender fatos,
 - **melhorar habilidades motoras/cognitivas** por meio de **práticas** e
 - **organizar conhecimento novo** em representações apropriadas para uso posterior.

Conceito de Aprendizagem de Máquina

O que é aprendizagem de Máquina ?



Conceito de Aprendizagem de Máquina

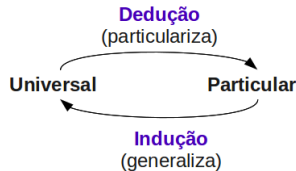


- **Capacidade de melhorar o desempenho** na realização de alguma tarefa por meio da experiência. (Mitchell, 1997)
- **Aprendizado é qualquer mudança em um sistema que melhore o seu desempenho na segunda vez** que ele repetir a mesma tarefa ou outra tarefa tirada da mesma população (Simon, 1983)
- **Processo de indução de uma hipótese a partir da experiência.**(Facelli e outros,2011).

Aprendizagem indutiva

- Considere um **conjunto de dados** de um hospital, por exemplo.
 - **Cada dado** (objeto, exemplo, padrão ou registro)
 - **corresponde a um individuo do dominio**. Ex: paciente
 - **é formado por uma tupla contendo características** (campo, atributos) do paciente.
 - **Atributo de entrada (atributo previsor)**. Ex: idade, sexo, sintomas e resultados de exames
 - **Atributo de saída (atributo alvo ou meta)**. Ex: diagnóstico, doença

Aprendizagem Indutiva



- **Indução da Hipótese:**

- O objetivo de um algoritmo de aprendizagem de máquina é **aprender, a partir de um subconjunto dos dados**, dito conjunto de treino, **um modelo ou hipótese** capaz de relacionar os valores dos atributos de um dado do conjunto de treino ao valor de seu atributo de saída.
- **Aprendizagem indutiva : generalização a partir de um conjunto de exemplos.**

Aprendizagem Indutiva

- **Processo de mapear atributos de entrada aos de saída** (aproximação de uma função ou modelo. A função ou modelo é a hipótese).
- **Formalmente:**
 - Dizemos que um **exemplo é um par** $(x, f(x))$, onde x é a **entrada** e $f(x)$ é a **saída da função aplicada a x** . A tarefa da **indução** é:
 - Dada **uma coleção de exemplos de f** , retornar **uma função h que se aproxime de f** .
 - A função h é **chamada de hipótese**.

Espaço de Hipóteses

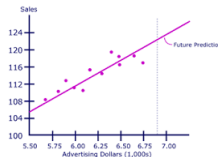
- Espaços de busca em problemas de aprendizagem são grandes.
 - **Dificuldade:** escolher entre diferentes generalizações suportadas pelos dados de treino.
 - **Objetivo: hipóteses consistentes e simples.**
 - **Consistente:** que concorda com todos os dados.
 - **Simples:** hipóteses que não são mais simples que os próprios dados estão deixando de extrair algum padrão dos dados.

Paradigmas e Tarefas de Aprendizagem

- **Paradigma de aprendizagem** é definido pela natureza do problema. Tipo de realimentação usada pelo algoritmo para aprender.
 - Podem ser:
 - **Supervisionado**: aprendizagem de uma função h a partir de exemplos (amostras rotuladas), de entradas (x) e saídas correspondentes ($f(x)$). Com crítica referente ao erro da saída.
 - **Não-supervisionado**: aprendizagem a partir de as amostras não são rotuladas. Essa abordagem não usa os atributos de saída. Sem crítica, usa regularidades e propriedades estatísticas dos dados.
 - **Por reforço**: processo de aprendizagem baseado em punição e recompensa. Reforça uma ação positiva e penaliza uma negativa. Crítica apenas de desempenho.

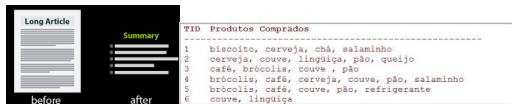
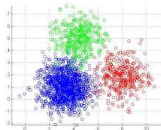
Paradigmas e Tarefas de Aprendizagem

- As **tarefas de aprendizagem** podem ser: **preditivas** ou **descritivas**
 - **preditivas**: tarefa supervisionada, sua meta é encontrar uma função (modelo ou hipótese) a partir dos dados de treino que possa ser usada para prever um rótulo (classe) ou valor de um novo exemplo.
 - Ex: **classificação** (rótulos discretos), **regressão** (rótulos contínuos)



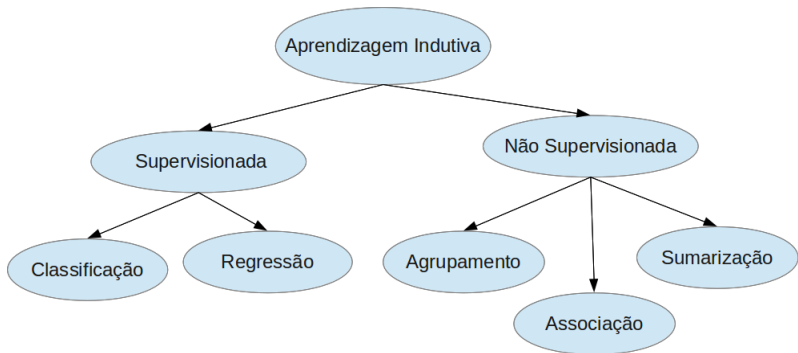
Paradigmas e Tarefas de Aprendizagem

- As tarefas de aprendizagem podem ser: **preditivas** ou **descritivas**
 - **descritivas**: tarefa não supervisionada, sua meta é explorar ou descrever um conjunto de dados. (não usam atributos de saída)
 - Ex: **agrupamento** (divisão em grupos baseada em similaridade), **sumarização** (descrição simples e compacta), **associação** (relações frequentes entre dados)



Paradigmas e Tarefas de Aprendizagem

- Resumo:



Abordagens Usadas

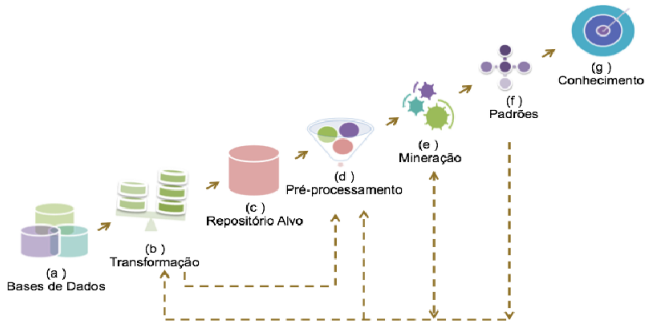
- A abordagem (método) usado para aprender podem ser:
 - **Simbólica**: a capacidade de descrever os padrões extraídos em uma linguagem compreensível para os usuários. Usa descrições simbólicas. Ex: árvores de decisão
 - **Conexionista**: inspiradas no modelo biológico do sistema nervoso. Ex: redes neurais:
 - **Estatística**: usa modelos estatísticos para encontrar uma boa aproximação da hipótese. Ex: redes bayesianas
 - **Evolutiva**: baseada na teoria da evolução de Darwin. Ex: algoritmos genéticos

Algumas Aplicações

- Reconhecimento de fala
- Predição de taxas de cura de pacientes em diferentes doenças
- Detecção de fraudes em cartões de crédito;
- Condução autônoma de carros;
- Desempenho em jogos complexos semelhante a campeões.
- Diagnóstico de doenças graves por análise de dados.

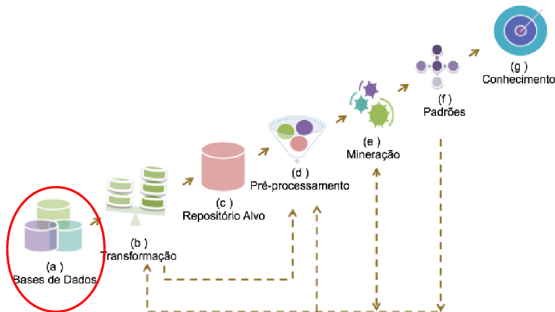
Processo de Descoberta de Conhecimento

- **Knowledge Discovery in Databases (KDD):** consiste em uma série de passos bem definida cujo meta é transformar dados em conhecimento.



Processo de Descoberta de Conhecimento

- Knowledge Discovery in Databases (KDD):
 - (a) Bases de Dados:
 - Bases de dados distintas, disponíveis em diferentes ambientes e em diferentes formatos
 - Bases heterogêneas

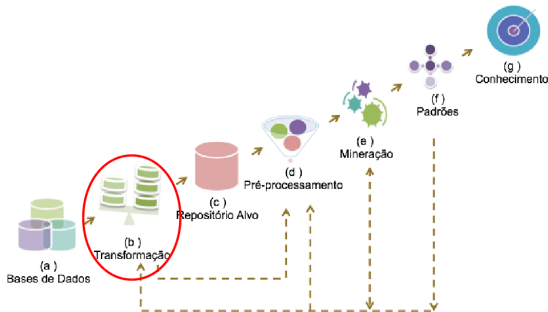


Processo de Descoberta de Conhecimento

- Knowledge Discovery in Databases (KDD):

- (b) Transformação :

- **Limpeza** dos dados brutos
- Etapa de **homogeneização** dos dados

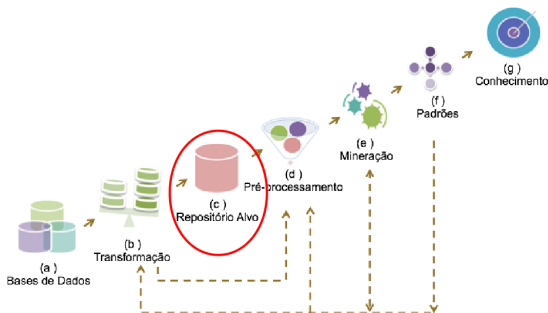


Processo de Descoberta de Conhecimento

- Knowledge Discovery in Databases (KDD):

- (c) Repositório Alvo :

- Histórico de registros
- Dados homogêneos e integrados

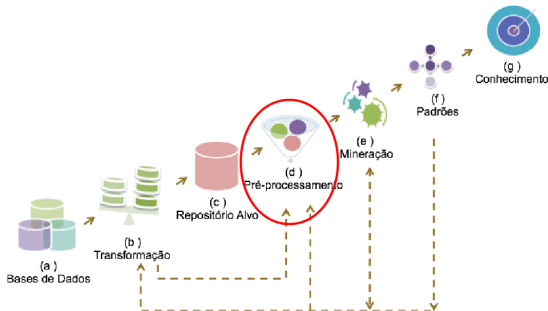


Processo de Descoberta de Conhecimento

- Knowledge Discovery in Databases (KDD):

- (d) Pré-processamento :

- Etapa de **ajuste fino dos dados** para atender ao objetivo da aprendizagem de máquina
- **~85% de todo o processo** (trabalhosa, mas valiosa)

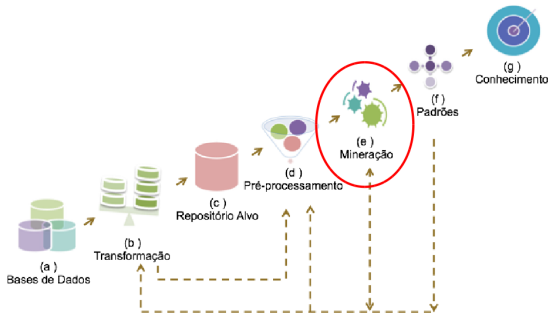


Processo de Descoberta de Conhecimento

- Knowledge Discovery in Databases (KDD):

- (e) Mineração :

- Usa Algoritmos de aprendizado de máquina
- Análise de uma séries de dados para compreensão do domínio
- Resultados compreensíveis e especialmente úteis

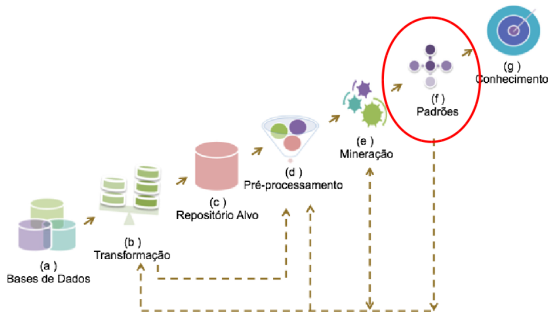


Processo de Descoberta de Conhecimento

- Knowledge Discovery in Databases (KDD):

- (f) Padrões :

- Modelos induzidos
- Modelos que indicam relações entre dados



Processo de Descoberta de Conhecimento

- Knowledge Discovery in Databases (KDD):

- (g) Conhecimento :

- Abstração dos modelos para uma real aplicação e, com ela, atingir melhores resultados

