

Aprendizado de Máquina

Aula 14: AutoML

André C. P. L. F de Carvalho
ICMC/USP

andre@icmc.usp.br



Tópicos a serem abordados

- Como selecionar o algoritmo mais adequado para uma dada tarefa
- Algoritmo mestre
- AutoML
- Meta-aprendizado
- Otimização
- Híbrida
- Meta-atributos

Tarefas preditivas

x_{11}	x_{12}	...	x_{1m}	y_1
x_{21}	x_{22}	...	x_{2m}	y_2
\vdots	\vdots		\vdots	\vdots
x_{n1}	x_{n2}	...	x_{nm}	y_n

Conjunto de treinamento

Algoritmo de AM

Treinamento

$f(x)$

Modelo de Classificação
/ Regressão

Indução

$f(\text{ })$

$x_1 \ x_2 \ \dots \ x_m$

Exemplo de teste

Classe /
Número

Teste

Dedução

Aprendizado de máquina

- Existem dezenas de milhares de algoritmos de aprendizado de máquina
 - Além disso, centenas de novos são propostos a cada ano
 - Levando em conta novos aspectos
 - Usando novas abordagens ou alterando abordagens existentes
 - Gerais ou adaptados para domínios específicos de aplicações ou problemas teóricos

Questão chave

- Como ter o melhor desempenho para uma nova tarefa de aplicação de aprendizado de máquina?
 - Qual algoritmo aprendizado de máquina pode induzir o melhor modelo para um novo conjunto de dados?

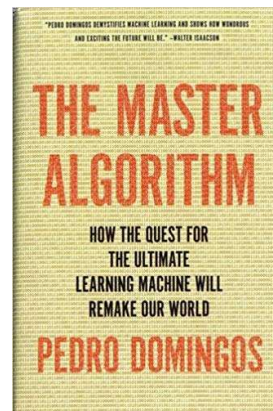
Questão chave

- Como ter o melhor desempenho para uma nova tarefa de aplicação de aprendizado de máquina?
 - Qual algoritmo aprendizado de máquina pode induzir o melhor modelo para um novo conjunto de dados?
 - Duas hipóteses:
 - Há um algoritmo mestre
 - O algoritmo mais adequado é o domínio (dados) dependente

Algoritmo Mestre (Master Algorithm)

Existe um algoritmo que pode superar todos os outros em qualquer tarefa de análise de dados

Pedro Domingos, University of Washington



“Panacea: Deusa da cura (mitologia grega)”

Problema da superioridade seletiva

- Cada algoritmo é melhor do que outros em um subconjunto de tarefas (Brodley 1995)
 - Não há almoço grátis (boca livre)
 - Cada algoritmo aprendizado de máquina tem um viés indutivo
 - Preferências de busca e representação
 - Necessário para que o aprendizado ocorra
- Hipótese: É possível selecionar o algoritmo mais apropriado para uma nova tarefa

É ainda mais complicado...

- Aplicação de aprendizado de máquina a um problema inclui mais do que apenas indução de modelo
- Limpeza de dados
- Pré-processamento de dados
- Ajuste de hiperparâmetros
- ...
- Pós-processamento
- Implementação e identificação de bugs

É ainda mais complicado

- Aplicação de aprendizado de máquina a um problema inclui mais do que apenas indução de modelo
- Limpeza de dados
- Pré-processamento de dados
- Ajuste de hiperparâmetros
- ...
- Pós-processamento
- Implementação e identificação de bugs

Aprendizado de Máquina
de ponta-a-ponta

Aprendizado de máquina de ponta-a-ponta

Inclui vários aspectos

Lidar com
valores ausentes

Lidar com dados
desbalanceados

Extrair atributos

Selecionar
atributos



Escolher/Modificar
algoritmo de AM

Ajustar
hiperparâmetros

Verificar overfitting

Descobrir bugs

Aprendizado de máquina de ponta-a-ponta

Inclui vários aspectos interdependentes

Lidar com valores ausentes

Lidar com dados desbalanceados

Extrair atributos

Selecionar atributos



Adaptado de Rick Caruana, Research opportunities in AutoML Microsoft Research

Escolher/Modificar algoritmo de AM

Ajustar hiperparâmetros

Verificar overfitting

Descobrir bugs

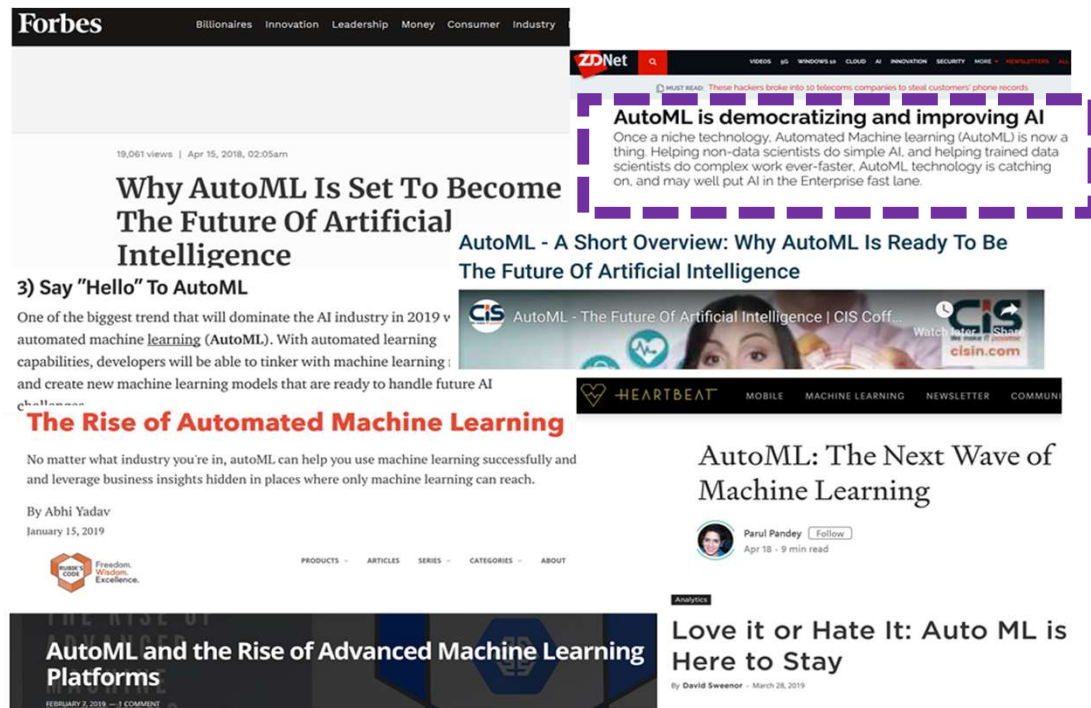
Questão chave revisitada

- Como ter o melhor desempenho para uma nova tarefa de aplicação aprendizado de máquina?
 - Qual algoritmo aprendizado de máquina pode induzir o melhor modelo para um novo conjunto de dados?
 - Quais são as melhores técnicas de pré-processamento?
 - Quais são os melhores valores para os hiperparâmetros?
 - ...
 - Qual é o melhor pipeline experimental?

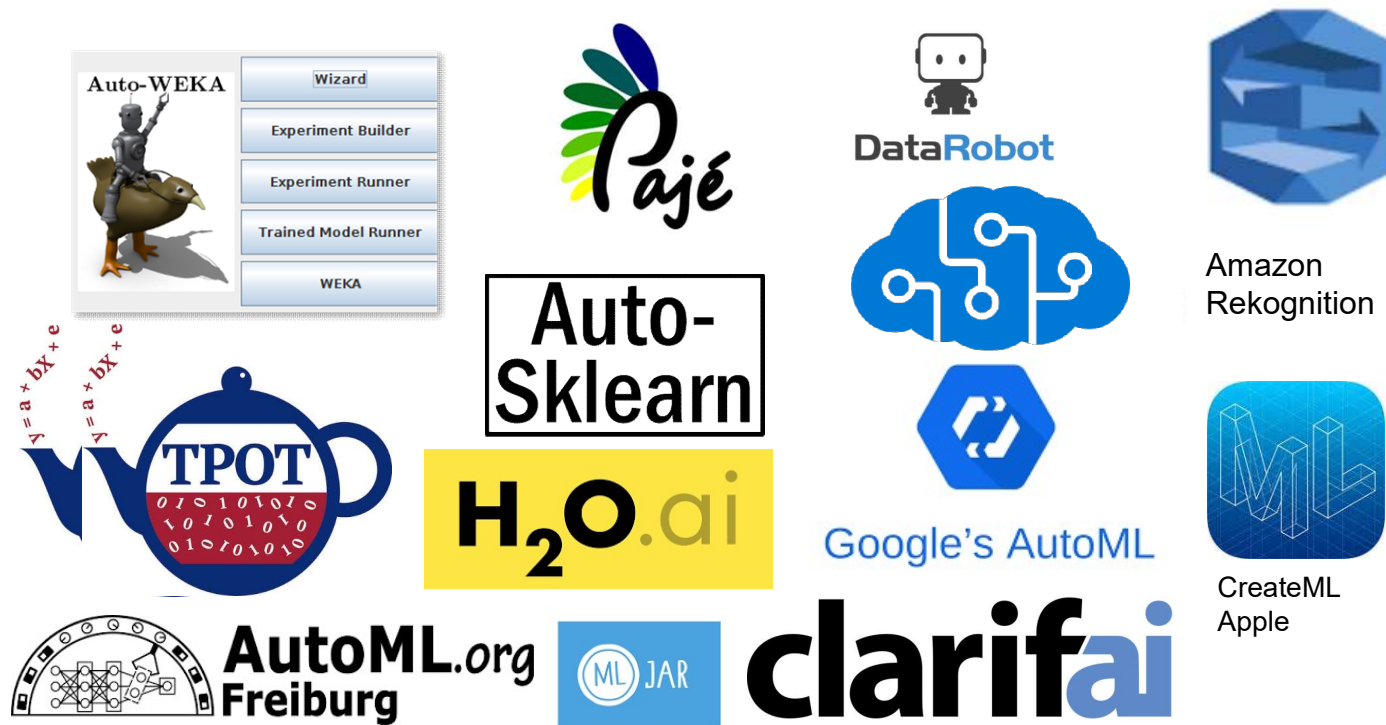
A solução mais apropriada

- Nova hipótese:
 - É possível selecionar não apenas o algoritmo mais adequado para uma nova tarefa, mas também
 - Técnicas de pré-processamento (pós-processamento)
 - Valores para os hiperparâmetros
 - ...
- Aprendizado de máquina automático (automatizado) – AutoML

AutoML



Ferramentas de AutoML



Ferramentas de AutoML



AutoML

- Automatiza aplicação de aprendizado de máquina a problemas reais
 - Apoia tanto leigo quanto especialista
- Engloba vários tópicos:
 - Otimização Bayesiana
 - Otimização combinatória
 - Aprendizado de máquina
 - Meta-aprendizado
 - Transferência de aprendizado

Principais abordagens de AutoML

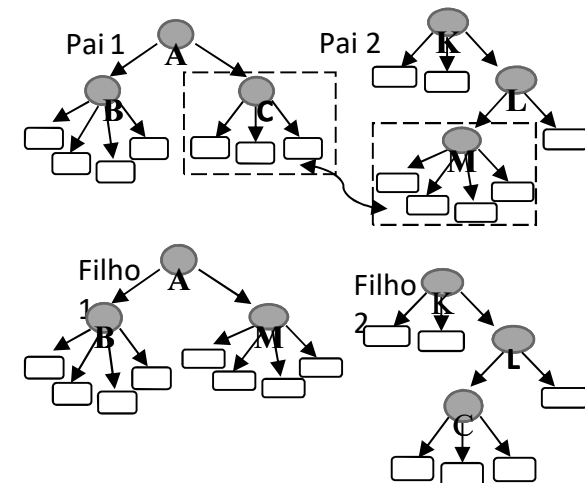
- Otimização
 - Algoritmos e/ou hiperparâmetros
 - Propõe o que pode não existir
- Meta-aprendizado (MtL)
 - Algoritmos e/ou hiperparâmetros
 - Seleciona entre o que já existe
- Híbrido
 - Combina abordagens anteriores

Otimização

- Ajuste de hiperparâmetros
 - Redes neurais artificiais
 - Máquinas de vetores de suporte
- Ajustar modelos
 - Árvores de decisão
- Projeto de novos algoritmos
 - Algoritmos de aprendizado de conjuntos de regras
 - Algoritmos de indução de árvores de decisão
 - Algoritmos de classificação bayesiana

Ajuste de modelos

- Buscar pela melhor árvores de decisão para um conjunto de dados de treinamento
 - Utilizar técnicas de otimização para construir, a partir de um conjunto de árvores, uma árvore melhor
 - Desempenho preditivo
 - Interpretação
 - Árvores podem ser geradas pelo mesmo algoritmo ou por diferentes algoritmos de indução de árvores de decisão
 - Ex.: CART, C4.5



Projeto de algoritmos: algoritmo HEAD-DT

- Algoritmo baseado em heurísticas evolutivas
 - Automatiza o projeto de novos algoritmos de indução de árvores de decisão
 - Diferente do ajuste de modelos (árvores de decisão)
- HEAD-DT pode projetar novos algoritmos em segundos
 - Algoritmos de indução de árvores de decisão levam meses ou anos para serem projetados por especialistas em aprendizado de máquina
 - Combina componentes de algoritmos existentes usando computação evolutiva
 - Algoritmos Genéticos (GA)
 - Programação Genética (GP)

Exemplo de algoritmo gerado

Algorithm

1. Recursively split nodes using the **Chandra-Varghese criterion**
2. Aggregate nominal splits in **binary subsets**
3. Perform step 1 until **class-homogeneity** or **the minimum number** of **5 instances** is reached
4. Perform **MEP pruning** with **m = 10**
5. When dealing with missing values:
Calculate the split of missing values by performing **unsupervised imputation**
Distribute missing values by **assigning the instance to all partitions**

Meta-aprendizado

- Aprenda com experiências de aprendizado
 - Aprende uma função (meta-modelo) associando:

Entrada Características extraídas de um conjunto de dados

Saída Recomendação de um ou mais algoritmos de aprendizado de máquina

- Meta-modelo pode
 - Prever os melhores algoritmos para novos conjuntos de dados
 - Fazer parte de um sistema de recomendação
- Aprendizado de nível básico e de nível meta



Meta-aprendizado

- Semelhante à aplicação convencional de um algoritmo de AM
 - Algoritmo de AM induz um modelo preditivo a partir de um conjunto de dados
 - Meta conjunto de dados (meta-dados)
 - Modelo induzido pode ser usado para prever resposta para dados novos
 - Recomenda técnicas para novos conjuntos de dados
 - Níveis de aprendizado base e meta

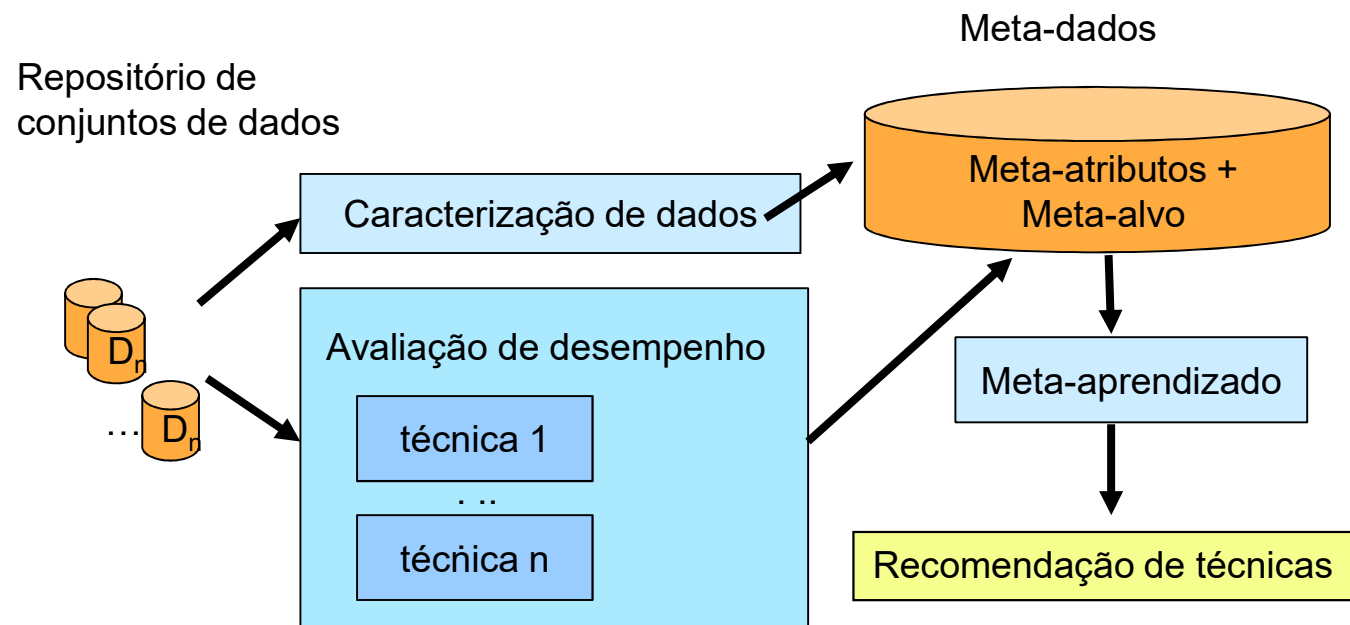
Tabela atributo-valor

Atributos de entrada (preditivos)				
	Altura	Tamanho Rabo	Textura	Classe
Exemplos (objetos, instâncias)	500	110	Manchas	Girafa
	440	90	Manchas	Girafa
	240	45	Listas	Zebra
	520	115	Manchas	Girafa
	260	50	Listas	Zebra
	230	50	Listas	Zebra
				Atributo alvo

Tabela para meta-aprendizado

Meta-atributos preditivos					
MAP1 MAP2 MAP3 MAP4					Algoritmo
Meta-exemplos (meta-objetos, meta-instâncias)	0.4	6	0.2	0.8	A
	0.1	2	0.2	0.5	A
	0.7	0	0.9	0.8	B
	0.2	4	0.7	0.1	A
	0.6	2	0.3	0.4	B
	0.1	7	0.1	0.9	B
Meta-atributo alvo					

Recomendação de técnicas

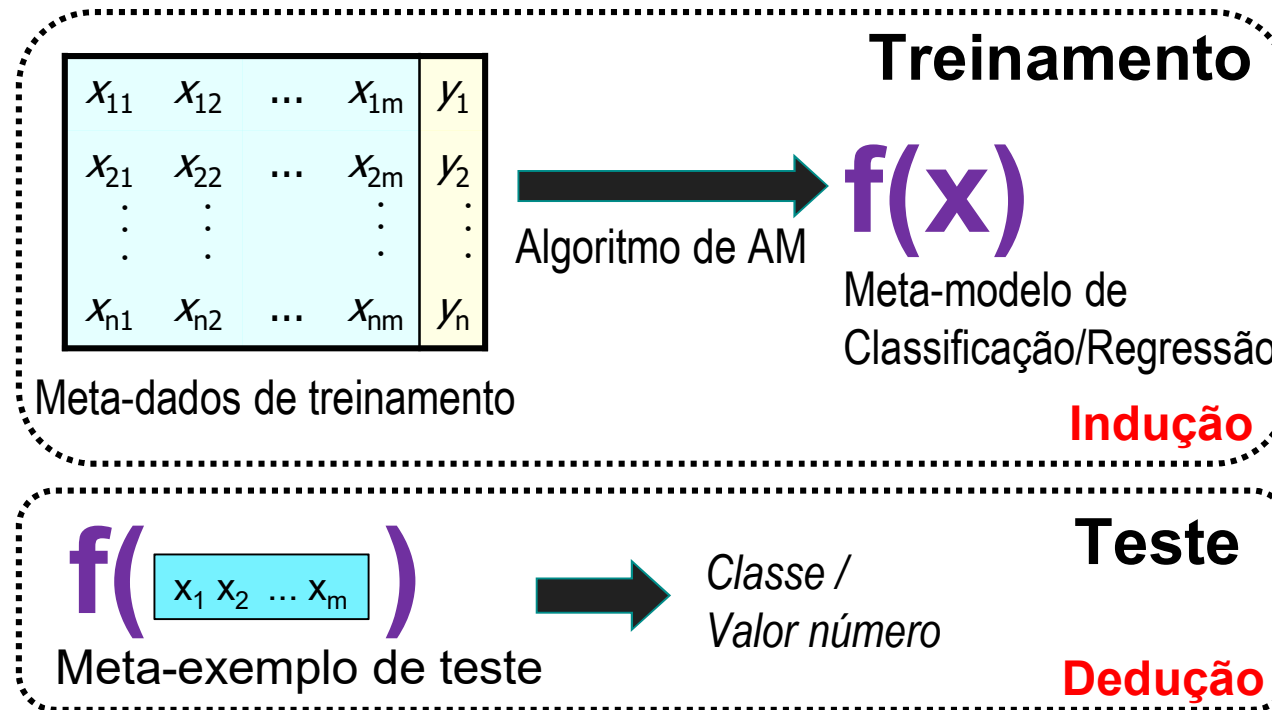


Adapted from P. Brazdil, C. Giraud Carrier, C. Soares and R. Vilalta, *Metalearning: Applications to Data Mining*, Springer

Indução de meta-modelos

- Aplicação convencional de AM
 - Aprende relação implícita entre meta-atributos e meta-alvo
 - Induz um meta-modelo preditivo
 - Atributos preditivos: meta-atributos
 - Atributo alvo: meta-alvo (desempenho de algoritmos de AM)
 - Regressão
 - Classificação

Indução e uso de meta-modelos



Geração de meta-dados

- Meta-exemplos
 - Atributo alvo (meta-atributo alvo, meta-alvo)
 - Desempenho de um conjunto de algoritmos (validação)
 - Melhor(es) algoritmo(s)
 - Atributos preditivos (meta-atributos preditivos)
 - Características do conjunto de dados
 - Caracterização direta
 - Baseada em modelos
 - Landmarking

Caracterização direta

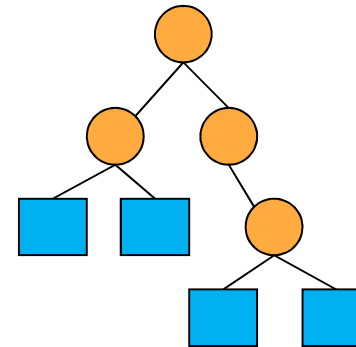
- Selecionar as descrições diretamente de cada conjunto de dados
 - Descrever os principais aspectos dos conjuntos de dados
- Meta-atributos
 - Medidas gerais
 - Medidas baseadas em estatística
 - Medidas baseadas em teoria da informação

Caracterização direta

- Exemplos de meta-atributos:
 - Número de classes
 - Número de atributos
 - #exemplos / #atributos
 - Correlação entre atributos preditivos
 - Correlação entre atributos preditivos e atributo alvo
 - Média da entropia das classes

Caracterização baseada em modelo

- Caracteriza um conjunto de dados pelas propriedades do modelo induzido
- Exemplos de meta-atributos:
 - Propriedades de uma AD induzida por um algoritmo de AM para um conjunto de dados
 - Número de nós folha
 - Formato da árvore
 - Profundidade da árvore
 - Largura da árvore
 - Grau de balanceamento da árvore



Landmarking

- Informação obtida ao executar um conjunto de algoritmos simples e rápidos (landmarkers)
 - Executar landmarks por um curto período
 - Landmarks devem ter diferentes vieses
 - Desempenhos dos algoritmos caracterizam um conjunto de dados
 - Conjuntos são semelhantes quando landmarks apresentam desempenhos semelhantes

Landmarking

- Exemplos de meta-atributos:
 - Revocação para algoritmo *landmark 1*
 - Precisão para algoritmo *landmark 1*
 - AUC para algoritmo *landmark 1*
 - Revocação para algoritmo *landmark 2*
 - Precisão para algoritmo *landmark 2*
 - AUC para algoritmo *landmark 2*
 - ...

Atributo alvo

- Medidas de desempenho
 - Desempenho preditivo
 - Acurácia, AUC, medida-F, MSE, ...
 - Custo de processamento
 - Tempo (aprendizado / uso)
 - Custo de armazenamento do modelo
 - Complexidade do modelo
 - Conhecimento (interpretabilidade)
 - Multiobjectivo

Formas de recomendação

- Melhor algoritmo
 - Se não estiver disponível ou não puder ser usado?
- Bons (estatisticamente equivalentes) algoritmos
 - Recomendação mais flexível
 - Permite selecionar algoritmo de acordo com preferências
- Ranking dos N melhores algoritmos
 - De acordo com uma medida de avaliação

Fim do
módulo