

Aprendizagem Incremental usando Tabelas Decisão Adaptativas

R. L. Stange, J. J. Neto

Resumo — A aprendizagem incremental requer que o mecanismo de aprendizagem seja capaz de acumular experiência dinamicamente. Os dispositivos adaptativos que possuem a capacidade de reter em suas regras informações extraídas de suas entradas, podem acumular essas informações e utilizá-las quando for necessário. Este trabalho propõe a utilização de dispositivos adaptativos para representar o conhecimento adquirido através da aprendizagem incremental. Alguns dos problemas do processo de aprendizagem em classificadores são apresentados. Um exemplo ilustrativo utilizando tabelas de decisão adaptativa, como forma de representar o conhecimento, é mostrado.

Palavras chaves — Aprendizagem de Incremental, Adaptatividade, Tabela de Decisão Adaptativa, Representação do conhecimento.

I. INTRODUÇÃO

A APRENDIZAGEM incremental consiste de técnicas para fornecer ao mecanismo de aprendizagem a capacidade de extrair informações adicionais a partir dos dados, sempre que surgirem mais dados, ou quando se julgar conveniente. Além disso, o conhecimento adquirido anteriormente é preservado e existe a capacidade de incorporar novos conhecimentos, possivelmente introduzidos por novos dados [1].

Em muitas aplicações, novas informações podem surgir ao longo do tempo. Por exemplo, quando um sistema de reconhecimento facial é construído é difícil recolher todas as variações de imagens de face com antecedência, pois rostos são facilmente alterados. Neste caso, pode existir a necessidade de implementar processos dinâmicos baseados em aprendizagem incremental.

As técnicas de aprendizagem incremental envolvem a adaptação gradual das estruturas de aprendizagem.

A adaptatividade é a capacidade que um dispositivo adaptativo tem de modificar a sua estrutura em resposta ao seu histórico de operações e aos dados de entrada [7].

A tecnologia adaptativa é o conjunto de ferramentas, métodos e técnicas que permitem modelar problemas práticos utilizando adaptatividade.

A tecnologia adaptativa pode ser uma solução alternativa para os problemas de aprendizagem computacional. Com a incorporação da adaptatividade é possível capturar um aspecto fundamental da aprendizagem, a adaptação dinâmica das regras de aprendizagem em função de sua interação com o ambiente [8].

O objetivo deste trabalho é apresentar uma proposta de utilização de adaptatividade em mecanismos de aprendizagem incremental. As tabelas de decisão adaptativas são utilizadas como forma de representação do conhecimento adquirido durante o processo de aprendizagem incremental.

II. CONCEITOS

A aprendizagem de máquina está relacionada à correta manipulação de conhecimento prévio e a novas observações que possam levar a novos conhecimentos [10]. A manipulação de conhecimento está relacionada aos métodos de inferência. Em particular, um método de inferência lógica é a aprendizagem indutiva.

A indução é uma forma de inferência lógica que permite obter conclusões genéricas a partir de um conjunto particular de exemplos [16].

A aprendizagem indutiva pode ser dividida em aprendizagem supervisionada e não supervisionada [18].

Na aprendizagem supervisionada é fornecido ao algoritmo de aprendizado, ou indutor, um conjunto de exemplos de treinamento, para os quais o rótulo da classe é conhecido. Em geral, cada exemplo de treinamento é descrito por um conjunto de características e pelo rótulo da classe que o exemplo de treinamento pertence. Quando o rótulo da classe é uma valor discreto, esse problema é conhecido como classificação. Quando o rótulo da classe é um valor contínuo, o problema é denominado regressão [2].

Na aprendizagem não supervisionada, o algoritmo de aprendizado analisa um conjunto de exemplos, não rotulados, e forma agrupamentos ou *clusters* desse conjunto de exemplos.

Neste trabalho é considerada a aprendizagem supervisionada.

R. L. Stange é estudante de Mestrado em Engenharia Elétrica da Escola Politécnica da Universidade de São Paulo, Departamento de Engenharia de Computação e Sistemas Digitais, Av. Prof. Luciano Gualberto, travessa 3, número 158 - Cidade Universitária - São Paulo - SP - CEP: 05508-900 (e-mail: rlstange@usp.br).

J. J. Neto atua no Departamento de Engenharia de Computação e Sistemas Digitais (PCS) da Escola Politécnica (POLI) da Universidade de São Paulo (USP), Av. Prof. Luciano Gualberto, Trav. 3, N. 158 - CEP: 05508-900 - São Paulo/SP, Brasil. (e-mail: joao.jose@poli.usp.br).

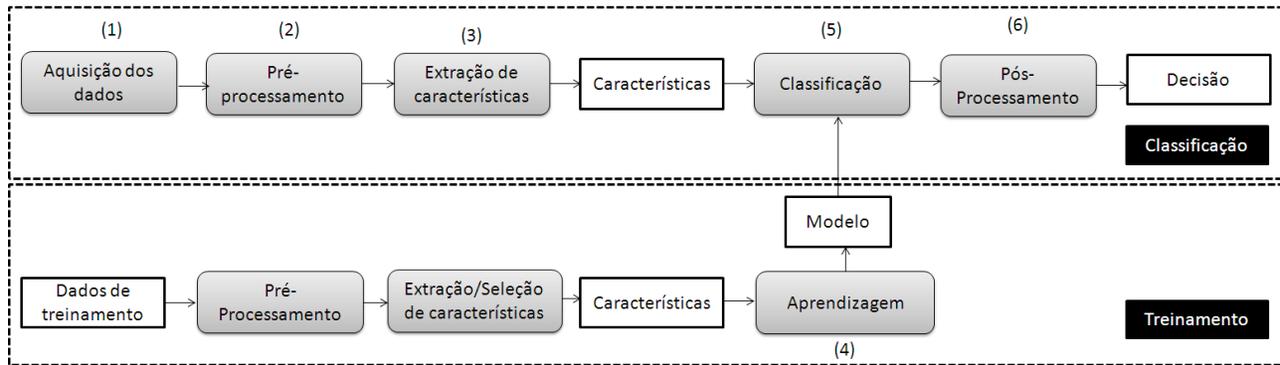


Figura 1. Processo de reconhecimento de padrão (Adaptada de [2]).

Conforme a disponibilidade dos exemplos de treinamento, os algoritmos de aprendizagem podem ser classificados em incrementais e não incrementais [10].

Na aprendizagem não incremental é necessário que todos os exemplos de treinamento estejam disponíveis simultaneamente para a indução do modelo de aprendizagem do classificador. Algoritmos não incrementais são apropriados para problemas de classificação com dados de treinamentos estáticos, ou seja, quando não haverá mudanças no conjunto de treinamento.

Na aprendizagem incremental o modelo de aprendizagem é modificado, se necessário, quando novos dados são disponibilizados. A vantagem em utilizar algoritmos incrementais é que o conhecimento pode ser atualizado de forma flexível. Além disso, em alguns casos é mais eficiente revisar um modelo de aprendizagem do que construir um novo modelo, cada vez que um novo exemplo é disponibilizado [14].

O reconhecimento de padrão de acordo com [14] é a descoberta automática de regularidades em dados, através de algoritmos computacionais, e uso dessas regularidades para classificar os dados em categorias. O termo genérico “padrão” é utilizado para referir-se a esses dados ou objetos.

A Fig. 1 mostra o diagrama geral para o processo de reconhecimento de padrão na aprendizagem supervisionada. O processo de reconhecimento de padrão é dividido basicamente em treinamento e classificação. Na componente de treinamento a finalidade é inferir, a partir de um conjunto de exemplos de treinamento, um conjunto de regras que define um padrão. O resultado final do treinamento é um modelo que descreve os padrões. Na componente de classificação, o modelo é utilizado com a finalidade de classificar novos indivíduos de acordo com critérios definidos durante sua construção.

A seguir são descritas as etapas do processo de reconhecimento de padrão de acordo com [2], identificadas na Fig. 1.

(1) Aquisição de dados e sensoriamento: Nesta etapa são realizadas as medições das variáveis físicas dos indivíduos (eg.: câmera para capturar imagens, microfone para capturar som, etc.).

(2) Pré-processamento: A etapa de pré-processamento ocorre tanto na fase de treinamento quanto na fase de classificação. O objetivo é remover eventuais ruídos nos dados

(3)

ou realizar qualquer outro tratamento sobre os dados capturados, para simplificar as operações subsequentes (eg.: ajustar distorção de imagem). No pré-processamento pode ocorrer um processo denominado segmentação. Na segmentação, os padrões devem ser separados para evitar sobreposições.

(4) Extração e seleção de características: A seleção de características é a tarefa de identificar e selecionar um subconjunto de características discriminantes, ou correlacionadas. Na extração de características novas características são geradas a partir de transformações ou combinações de características do conjunto original.

(5) Aprendizagem do modelo: Algoritmos de aprendizagem mapeiam as características entre grupos de padrões.

(6) Classificação: A classificação utiliza as características e o modelo de aprendizagem para atribuir categorias ou classes aos novos indivíduos.

(7) Pós-processamento: Após a classificação, uma avaliação de confiança nas decisões tomada pelo classificador é realizada. Na etapa de pós-processamento também podem ocorrer combinações de classificadores.

III. TECNOLOGIA ADAPTATIVA

As pesquisas em adaptatividade investigam soluções para diversos problemas complexos de teoria da computação [5], de aprendizagem de máquina [8][9], de tomada de decisão [15], de engenharia da computação [8], entre outros.

A. Dispositivos Adaptativos

O modelo geral para um dispositivo adaptativo é definido como um conjunto finito de regras que pode sofrer modificações dinamicamente [6].

Um dispositivo dirigido por regras tem seu comportamento completo descrito na forma de um conjunto finito de regras. As regras usualmente são expressas como cláusulas condicionais do tipo “Se C_i então A_i ”.

O conjunto de regras $R = \{r_1, r_2, \dots, r_k\}$, onde k é o número de regras, determina o funcionamento do dispositivo adaptativo, o qual muda sucessivamente de configuração em resposta a estímulos externos de entrada. Cada regra r_i , para $0 < i \leq k$ pode ser representada no formato $r_i = (C_i, A_i)$, onde C é conjunto de condições $C = \{\text{condição}_1, \text{condição}_1, \dots, \text{condição}_n\}$, onde $n \geq 1$ é número de condições e A é o conjunto de ações $A = \{\text{ação}_1, \text{ação}_2, \dots, \text{ação}_m\}$, onde $m \geq 1$ é o número de ações.

Uma formulação para um dispositivo guiado por regras não adaptativo encontra-se resumida na Tabela 1, e sua a formulação completa, em [6].

Um dispositivo adaptativo AD é dito com adaptatividade básica quando reduzido a um dispositivo não adaptativo subjacente ND, acoplado a um mecanismo adaptativo AM.

O mecanismo adaptativo incorpora um conjunto de funções adaptativas, as quais são acopladas às regras do dispositivo subjacente e, são capazes de alterar convenientemente o seu próprio conjunto das regras provocando uma mudança de comportamento no dispositivo AD.

TABELA 1
FORMULAÇÃO DE UM DISPOSITIVO GUIADO POR REGRAS.

FORMULAÇÃO - DISPOSITIVO NÃO ADAPTATIVO	
ND = (C, R, S, c ₀ , A)	
C	Conjunto de todas as possíveis configurações do dispositivo;
R	Relação de mudança de configuração, $R \subseteq C \times (S \cup \{\varepsilon\}) \times C$.
S	Conjunto de estímulos de entrada do dispositivo;
c ₀	Configuração inicial única do dispositivo, pertence ao conjunto C
A	Conjunto das configurações de aceitação do dispositivo;

Um dispositivo AD pode ser descrito como uma sêxtupla, conforme a formulação apresentada na Tabela 2.

As funções adaptativas são descritas basicamente por um cabeçalho e um corpo, similar às funções em linguagens de programação, conforme Tabela 3.

As funções adaptativas, em seu núcleo, referem-se a operações básicas de edição do conjunto de regras que definem o dispositivo adaptativo, representadas pelas ações adaptativas elementares de consulta, remoção e inserção de regra. As ações adaptativas elementares são apresentadas na Tabela 4.

Têm sido alvos frequentes de pesquisa os seguintes dispositivos adaptativos: os autômatos adaptativos [7], as gramáticas adaptativas [4], as árvores de decisão adaptativas [11] e as tabelas de decisão adaptativa [6] [15].

TABELA 2
FORMULAÇÃO DE UM DISPOSITIVO ADAPTATIVO.

FORMULAÇÃO - DISPOSITIVO ADAPTATIVO BÁSICO	
AD = (C, RA, S, AA, c ₀ , RA ₀ , A)	
C	Conjunto de todas as possíveis configurações do dispositivo adaptativo;
RA	Conjunto de todas as regras adaptativas do dispositivo adaptativo;
S	Conjunto de estímulos de entrada do dispositivo adaptativo;
AA	Conjunto de funções adaptativas φ , incluindo a função nula λ ;
c ₀	Configuração inicial $c_0 \in C$ do dispositivo adaptativo, que deve ser única;
RA ₀	Conjunto inicial de regras adaptativas do dispositivo adaptativo;
A	Conjunto de todas as configurações de aceitação do dispositivo adaptativo;

TABELA 3
DECLARAÇÃO DAS FUNÇÕES ADAPTATIVAS.

DECLARAÇÃO - FUNÇÃO ADAPTATIVA φ	
CABEÇALHO	
Nome φ	Uma função adaptativa é referenciada por um nome;
Parâmetros	Ênupla ordenada $(\rho_1, \rho_2, \dots, \rho_p)$, de $p \geq 0$ parâmetros formais;
Variáveis	Um conjunto $\{v_1, v_2, \dots, v_v\}$, de $v \geq 0$ variáveis. Os valores das variáveis são preenchidos uma única vez pelas ações adaptativas de consulta;
Geradores	Um conjunto $\{\chi_1, \chi_2, \dots, \chi_g\}$, de $g \geq 0$ geradores. Os valores dos geradores são preenchidos com novos valores a cada chamada da função adaptativa.
CORPO ($\beta\mu, \Delta, \alpha\nu$)	
Função adaptativa anterior	Chamada $\beta\mu$ de uma função adaptativa anterior (μ) executada antes da aplicação das ações adaptativas elementares definidas no conjunto Δ .
Função adaptativa posterior	Chamada $\alpha\nu$ de uma função adaptativa posterior (ν) executada após a aplicação das ações adaptativas elementares definidas no conjunto Δ .
Núcleo da função adaptativa	Núcleo da função adaptativa φ consiste do conjunto $\Delta = \{\delta_1, \delta_2, \dots, \delta_e\}$, contendo $e \geq 0$ ações adaptativas elementares δ_i , para $0 \leq i < e$.

TABELA 4
AÇÕES ADAPTATIVAS ELEMENTARES.

AÇÕES ADAPTATIVAS ELEMENTARES δ_i	
$\delta_i = \otimes [r]$	
\otimes	Representa um operador aplicado à regra r e ao conjunto RA_t de regras que definem o comportamento do dispositivo adaptativo AD.
RA _t	Conjunto RA no passo t de operação do dispositivo AD.
r	Padrão para a regra sobre a qual deve incidir a ação adaptativa elementar.
OPERADORES \otimes	
Inserção +	Inclui no conjunto RA _t uma nova regra r.
Remoção -	Elimina do conjunto RA _t a regra r.
Consulta ?	Pesquisa em RA _t o padrão r e preenche variáveis com os resultados da busca.

B. Tabelas de Decisão Adaptativas

Nas tabelas de decisão adaptativas o dispositivo subjacente é uma tabela de decisão convencional. Uma tabela de decisão convencional é composta por colunas que representam conjuntos de regras associadas a condições e ações [3]. A primeira coluna, partindo da primeira linha da tabela, representa um conjunto de condições e, a seguir, encontra-se um conjunto de ações. A tabela de decisão opera verificando as condições de acordo com os valores definidos nas colunas de regras. Quando a condição é satisfeita de acordo com uma determinada regra, essa regra é considerada válida e todas as ações a ela associadas são executadas.

A formulação da tabela de decisão é apresentada na Tabela 5.

TABELA 5
FORMULAÇÃO DA TABELA DE DECISÃO CONVENCIONAL.

FORMULAÇÃO – TABELA DE DECISÃO CONVENCIONAL	
TD = (CT, R, CV, t ₀ , AT, A)	
CT	Conjunto de todas as configurações possíveis da tabela de decisão;
R	Conjunto finito de regras de decisão: $R = \{r_i, 1 \leq i \leq n\}$. Cada regra $r_i = (c_{ij}, a_{ik}) \in R$, onde: c_{ij} : representa um valor para a condição c_i na regra r_i ; a_{ik} e representa um valor para a alternativa ou ação a_k para a regra r_i ;
CV	Conjunto finito dos valores c_{ij} válidos para as condições C_i
t ₀	Configuração inicial da tabela de decisão;
AT	Conjunto de configurações aceitas da tabela de decisão;
A	Conjunto finito de ações;

Uma versão adaptativa de uma tabela de decisão convencional pode ser obtida adicionando-se a ela linhas, onde são incluídas as funções adaptativas. Além disso, a cada coluna que representa uma regra simples, deve ser adicionada uma chamada para uma função adaptativa associada à execução de uma regra em particular. Com isso, sempre que uma regra adaptativa é aplicada, uma ação adaptativa é invocada, permitindo que sejam feitas mudanças no conjunto de regras.

Uma tabela de decisão adaptativa ou TDA, cujo formalismo está apresentado em [6][15], é um dispositivo guiado por regras que permite modificações dinâmicas no conjunto de regras que compõe a tabela de decisões, através de ações adaptativas. As tabelas de decisão adaptativas são formadas por um conjunto de condições, ações, regras e funções adaptativas. A estrutura geral gráfica de uma tabela de decisão adaptativa é apresentada na Tabela 6, baseada no formato descrito em [6] e na tabela de decisão convencional descrita em [3].

Na Tabela 6, as linhas que compõem a tabela de decisão TD correspondem ao dispositivo não adaptativo subjacente ND. O mecanismo adaptativo é obtido acrescentando uma camada adaptativa no dispositivo ND. A camada adaptativa composta por um conjunto de linhas para a definição das funções adaptativas.

Os elementos que compõem a tabela de decisão adaptativa são descritos a seguir:

1. *Linha Cabeçalhos* → (Tags): especifica o cabeçalho das colunas. O título de cada coluna pode ser:

- “H” que representa o cabeçalho da função adaptativa;
- “?”, “-” ou “+” que indicam respectivamente uma ação elementar de consulta, remoção e inserção;
- “S”, “E” ou “R” que indicam respectivamente uma regra delimitadora inicial, final ou uma regra da tabela de decisão convencional.

2. *Linhas das Condições*: cada linha de condição (condição₁,..., condição_n) corresponde a um rótulo de condição;

3. *Valores das condições*: em suas células, são indicados os valores das condições (c_{11}, \dots, c_{kn}). Cada valor pode corresponder a entradas limitadas (“S”, “N” ou branco) ou a entradas estendidas;

4. *Linhas das Ações*: cada linha representa uma ação (ação₁,...,ação_m), que pode ser executado em resposta ao conjunto de condições;

5. *Valores das Ações*: devem ser marcadas as células das ações (a_{11}, \dots, a_{kn}) que serão executadas quando as regras são avaliadas;

6. *Funções adaptativas*: as linhas nome da função adaptativa (Nome ϕ), parâmetros (ρ_1, \dots, ρ_p), variáveis (v_1, \dots, v_v) e geradores (χ_1, \dots, χ_g) são definidas, convenientemente;

7. *Chamadas às ações adaptativas*: em suas células assinalam-se aquelas que serão chamadas antes ou depois da aplicação da regra.

Na execução de uma regra em uma TDA, primeiramente são verificadas as regras não adaptativas e, se uma única delas se aplica, as ações correspondentes são executadas. Caso mais de uma regra não adaptativa satisfaça a condição, as ações correspondentes às mesmas devem ser aplicadas em paralelo, como previamente definido para tabelas de decisão convencionais. Porém, se nenhuma regra não adaptativa satisfizer a condição, trata-se de uma condição não prevista e, no caso de uma tabela de decisão convencional, não haveria como prosseguir.

TABELA 6
TABELA DE DECISÃO ADAPTATIVA.

			r ₁	r ₂	...	r _k	
Cabeçalhos (Tags) →		(H,+, -,?)	S	R	R	E	
Tabela de decisão TD	Condições	condição ₁	c ₁₁	c ₂₁	...	c _{k1}	
		
		condição _n	c _{1n}	c _{2n}	...	c _{kn}	
	Ações	ação ₁	a ₁₁	a ₂₁	...	a _{k1}	
		
		ação _m	a _{1m}	a _{2m}	...	a _{km}	
Funções Adaptativas	Anterior	Nome ϕ	Declaração das funções				Chamadas às ações adaptativas
	Posterior	Nome ϕ					
	Parâmetros	ρ_1					
		...					
	Variáveis	v_1					
		...					
	Geradores	χ_1					
...							
	χ_g						

IV. TABELAS DE DECISÃO ADAPTATIVAS EM PROBLEMAS DE RECONHECIMENTO DE PADRÕES.

Neste trabalho, o modelo de aprendizagem é representado por um conjunto de regras do tipo “Se...então”. O conjunto de regras é utilizado para classificar novos indivíduos. A Fig. 2 mostra o processo de aprendizagem incremental baseado em regras, onde D_1, D_2, \dots, D_t são novos dados disponíveis em diferentes instantes no tempo. A base de regras R armazena as regras do modelo.

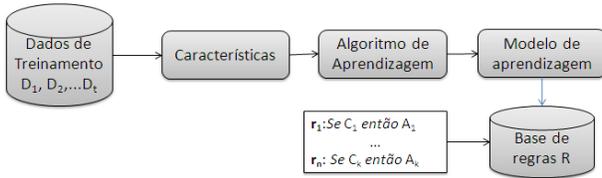


Figura 2. Modelo de aprendizagem incremental baseado em regras.

Na aprendizagem incremental utilizando adaptatividade, as regras de classificação podem ser alteradas, em resposta à estímulos de entrada, tais como quando novos exemplos de treinamento estão disponíveis. A adaptatividade é incorporada na base de regras através das regras adaptativas, que quando aplicadas alteram o modelo de aprendizagem ou base de regras.

A seguir, apresenta-se um estudo de caso, para explorar a utilização das tabelas de decisão adaptativas, na resolução de problemas de classificação.

Considere-se o conjunto de atributos $A = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7\}$, onde:

- $\alpha_1 =$ “temperatura corporal”, com temperatura corporal = {quente, frio};
- $\alpha_2 =$ “cobertura da pele”, com cobertura da pele = {cabelos, escamas, pelos, penas, espinhos, nada};
- $\alpha_3 =$ “ovíparo”, com ovíparo = {sim, não};
- $\alpha_4 =$ “criatura aquática”, com criatura aquática = {sim, não, semi};
- $\alpha_5 =$ “criatura aérea”, com criatura aérea = {sim, não};
- $\alpha_6 =$ “possui pernas”, com possui pernas = {sim, não};
- $\alpha_7 =$ “hiberna”, com hiberna = {sim, não}.

Considere-se o conjunto de classes distintas $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$, onde:

- $\omega_1 =$ “mamífero”;
- $\omega_2 =$ “pássaro”;
- $\omega_3 =$ “peixe”;
- $\omega_4 =$ “réptil”;
- $\omega_5 =$ “anfíbio”.

Dado o conjunto de dados de treinamento de vertebrados D com amostras rotuladas, conforme Tabela 7.

TABELA 7
DADOS DE TREINAMENTO [13]

Animal	Temperatura Corporal	Cobertura da pele	Ovíparo	Criatura Aquática	Criatura Aérea	Possui Pernas	Hiberna	Classe
Humano	Quente	Cabelos	Não	Não	Não	Sim	Não	Mamífero
Cobra	Fria	Escamas	Sim	Não	Não	Não	Sim	Réptil
Salmão	Fria	Escamas	Sim	Sim	Não	Não	Não	Peixe
Baleia	Quente	Cabelos	Não	Sim	Não	Não	Não	Mamífero
Sapo	Fria	Não possui	Sim	Semi	Não	Sim	Sim	Anfíbio
Morcego	Quente	Cabelo	Não	Não	Sim	Sim	Sim	Mamífero
Pomba	Quente	Penas	Sim	Não	Sim	Sim	Não	Pássaro
Gato	Quente	Pelos	Não	Não	Não	Sim	Não	Mamífero
Tartaruga	Fria	Escamas	Sim	Semi	Não	Sim	Não	Réptil
Pinguim	Quente	Penas	Sim	Semi	Não	Sim	Não	Pássaro
Porco Espinho	Quente	Espinhos	Não	Não	Não	Sim	Sim	Mamífero
Enguia	Fria	Escamas	Sim	Sim	Não	Não	Não	Peixe
Salamandra	Fria	Não possui	Sim	Semi	Não	Sim	Sim	Anfíbio

Considere que μ_j é um método de indução de árvore. Esse método é implementado pelo algoritmo ID3 [11]. A saída do algoritmo de aprendizagem é um classificador τ , neste caso,

representado por um dispositivo do tipo árvore de decisão. Dado um novo padrão χ , o classificador deve prever o valor correspondente da classe Ω associada ao padrão.

Com a finalidade de melhorar a legibilidade as informações contidas na árvore, são transformadas em um conjunto de regras do tipo “Se...então” e codificadas em uma tabela de decisão convencional para representar o classificador. O resultado é o conjunto de regras $R = \{r_1, r_2, \dots, r_k\}$, onde k é o número de regras. R é descrito na Tabela 8.

As tabelas de decisão, assim como as árvores de decisões ou redes neurais, são dispositivos capazes de representação do conhecimento para o reconhecimento de padrões.

Conforme a Fig. 1, na fase de avaliação do processo de aprendizagem, os dados de testes são submetidos ao classificador. Como resultado, quatro casos resultantes são possíveis.

TABELA 8
REGRAS DE CLASSIFICAÇÃO NA TABELA DE DECISÃO.

Regras →	r ₁	r ₂	r ₃	r ₄	r ₅	r ₆
Temperatura corporal	fria	-	-	-	-	-
Cobertura da pele	-	cabelo	escamas	-	Não possui	-
Ovíparo	-	-	-	sim	-	-
Criatura aquática	sim	não	não	não	-	-
Criatura aérea	-	-	-	-	-	-
Possui pernas	-	-	-	-	-	-
Hiberna	-	-	-	-	-	-
Classe	peixe	mamífero	réptil	pássaro	anfíbio	-

Caso 1: Apenas uma regra do conjunto R é aplicável ao dado de entrada e o classificador rotula o padrão corretamente. Por exemplo, para $\chi = \{fria, escamas, não, sim, não, não, não\}$ a regra aplicável é r_1 . Neste caso, o padrão é classificado como peixe. O exemplo χ é um salmão, portanto, a classificação está correta.

Caso 2: Nenhuma regra do conjunto R é aplicável. Isso significa que o classificador não é capaz de classificar χ . Os motivos podem ser diversos, tais como: χ não é um padrão, neste caso, χ deve ser descartado; χ é um padrão, porém o classificador não aprendeu as regras necessárias para classificá-lo, neste caso, este problema ocorreu na fase de aprendizagem e os motivos podem ser: o método de aprendizagem não é adequado, por exemplo, os atributos selecionados pelo método para discriminar as classes são insuficientes, ou não são bons discriminantes; ocorrência de problemas nas amostras de treinamento (dados ruidosos, dados ausentes, quantidade de dados insuficiente); χ é um novo padrão que deverá ser aprendido, *a priori*, não pertencia ao conjunto de treinamento D .

Caso 3: Existe mais de uma regra aplicável no conjunto R . Trata-se, portanto, de uma situação não determinística. Isso pode ocorrer quando os dados de treinamento são ambíguos ou quando o conjunto de treinamento é pequeno, entre outros motivos.

Caso 4: Existe uma regra aplicável no conjunto R, porém o classificador está rotulando os padrões de maneira incorreta. Isso pode ocorrer por várias razões, entre as quais: o método de aprendizagem é inadequado para as características dos dados de treinamento; os dados de treinamento são insuficientes para uma boa taxa de aprendizagem; ocorreu *overfitting*²[2].

A seguir, é ilustrado um exemplo de utilização das tabelas de decisão adaptativas para o Caso 2.

Considere $\chi = \{\text{quente, pelos, não, não, não, sim, não}\}$. Neste caso, não existe uma regra aplicável. Porém, é sabido que χ é um exemplo de gato e deveria ser classificado como mamífero. O algoritmo ID3 não foi capaz de aprender a regra que classifica um gato como mamífero.

Neste exemplo, utilizando uma tabela de decisão adaptativa uma regra adaptativa é adicionada ao conjunto de regras R da tabela de decisão. Essa regra é satisfeita toda vez que nenhuma regra não adaptativa é aplicável. A regra r_6 é uma regra adaptativa e possui a seguinte função adaptativa associada:

Exemplo do pseudocódigo para declaração de uma função adaptativa.

```
Função Adaptativa  $f_1$ (Parâmetros:  $p_1, p_2, p_3, p_4, p_5, p_6, p_7$ ) {
Variáveis:  $v_1, v_2, v_3, v_4, v_5, v_6, v_7$ 
Geradores: Geradores de rótulos  $*g_t$ 
Ações elementares  $\Delta$  {
 $\delta_1: ? [v_1, v_2, v_3, v_4, v_5, v_6, v_7]$  (Existe
uma regra na área temporária que
satisfaça as condições  $\chi$ ).
 $\delta_2: - [v_1, v_2, v_3, v_4, v_5, v_6, v_7]$  (Se
existir, remover).
 $\delta_3: + [p_1, p_2, p_3, p_4, p_5, p_6, p_7, *g_1]$  (Inse-
rir uma regra  $r_t$ )
.}
}
```

Para fins didáticos, o conjunto $\square\square$ é dividido em R_p , o conjunto de regras principais do dispositivo TDA e R_t , conjunto de regras temporárias do dispositivo TDA. Assim, $R = R_p \cup R_t$.

No caso de aplicação da regra r_6 em χ , a ação elementar de consulta irá pesquisar se existe uma regra em R_t que satisfaça χ . Se existir a ação elementar de remoção irá removê-la. Em seguida a ação elementar de inserção inclui uma regra r_{it} , para $i = 1, 2, \dots, n$, onde n é o número de regras temporárias. Por exemplo, ao receber as entradas $\chi_1 = \{\text{quente, pelos, não, não, não, sim, não}\}$, $\chi_2 = \{\text{quente, pelos, não, não, não, sim, não}\}$ e $\chi_3 = \{\text{quente, penas, sim, semi, não, sim, não}\}$, a regra r_6 é aplicada e o conjunto de regras R_t é modificado.

A Tabela 9 mostra a TDA após sucessivas transformações. As transformações do dispositivo são representadas por TDA_j

para $j = 0, 1, 2, \dots, n$, onde n é o número de modificações do dispositivo.

Posteriormente, as regras temporárias deverão ser utilizadas para inferir novas regras de classificação e assim aprender futuramente a classificar corretamente novos padrões não cobertos pelas regras iniciais.

É importante destacar que no exemplo da Tabela 2, apenas uma regra adaptativa foi inserida, porém é possível incluir outras regras adaptativas com funções adaptativas capazes de realizar diferentes modificações na tabela de decisão adaptativa.

V. CONSIDERAÇÕES FINAIS

A área de aprendizagem de máquina tem-se mostrado uma rica fonte de pesquisa para a exploração prática das aplicações dos fundamentos da tecnologia adaptativa.

Em problemas de classificação diversas soluções são propostas na tentativa de construir classificadores eficientes, na maioria dos casos, as melhorias ocorrem nas fases de treinamento e validação. Porém, muitos problemas só ocorrem na fase de classificação, quando os padrões a serem rotulados são submetidos ao classificador.

A aprendizagem incremental é uma solução desejável para adaptar gradualmente o modelo aprendido. A adaptatividade pode ser útil para detectar os problemas que podem ocorrer na fase de classificação e modificar o modelo, de maneira incremental. Embora existam maneiras de refazer a estrutura das tabelas de decisão após a obtenção de novos dados, em alguns casos isso pode ser ineficiente, podendo levar à perda, parcial ou total, de informações que haviam sido anteriormente aprendidas.

No contexto da aprendizagem incremental e mecanização da aprendizagem, o exemplo ilustrativo da seção IV mostra que a utilização de dispositivos adaptativos em classificadores é conveniente, pois sua operação poder ser descrita de forma incremental.

O equilíbrio entre a obtenção de modelos de aprendizagem compreensíveis e expressivos é adquirido com a utilização das tabelas de decisão adaptativas. A utilização da tabela de decisão como dispositivo subjacente para representar os modelos de classificação facilita o entendimento do modelo. Muitos modelos de classificação são como caixas pretas, as tabelas de decisão têm uma estrutura simples e compreensível, o que transforma os classificadores, vistos como caixas pretas, em ferramentas úteis para a descoberta do conhecimento.

A proposta do modelo de aprendizagem incremental baseado em regras adaptativas mostra as tabelas de decisão adaptativas como uma proposta atraente para mecanizar o processo de aprendizagem incremental em classificadores. Também permite estabelecer de forma satisfatória o domínio do problema, bem como identificar melhorias importantes para o processo de aprendizagem. Uma das vantagens da utilização dos dispositivos adaptativos é a facilidade de uso e simplicidade na aprendizagem incremental.

A grande vantagem da adaptatividade sobre muitas outras técnicas correntemente utilizadas para a formulação de modelos de representação e de manipulação do conhecimento

² *Overfitting*: aprendizagem superajustada ao conjunto de treinamento.

reside no fato de que: (a) a informação total, encerrada no dispositivo adaptativo, está representada integralmente no conjunto de regras; (b) a aprendizagem angariada em cada passo adaptativo do dispositivo encontra-se integralmente confinada à variação sofrida pelo conjunto de regras; (c) a observação, a identificação, a qualificação e a quantificação da evolução da aprendizagem, bem como a relação de causa e efeito entre a seqüência de entradas processadas, a das regras aplicadas e a das variações sofridas pelo conjunto de regras, tudo isso pode ser efetuado única e exclusivamente pela análise do conjunto de regras do dispositivo adaptativo e das suas variações ao longo da operação do dispositivo.

associadas às regras para tratar de outros problemas de classificação tais como dados de entradas ausentes ou com ruídos.

TABELA 9
TABELA DE DECISÃO ADAPTATIVA APÓS 3 MODIFICAÇÕES TDA₃.

Cabeçalhos (Tags)		H	?	-	+	r ₁	r ₂	r ₃	r ₄	r ₅	r ₆	r _{t1}	r _{t2}	r _{t3}
Condições	Temperatura corporal	p ₁	v ₁	v ₁	fria	-	-	-	-	-	-	quente	quente	quente
	Cobertura da pele	p ₂	v ₂	v ₂	-	cabelo	escamas	-	Não possui	-	-	pelos	espinhos	penas
	Oviparo	p ₃	v ₃	v ₃	-	-	-	sim	-	-	-	não	não	sim
	Criatura aérea	p ₄	v ₄	v ₄	-	-	-	-	-	-	-	não	não	semi
	Criatura aquática	p ₅	v ₅	v ₅	sim	não	não	não	-	-	-	não	não	não
	Possui pernas	p ₆	v ₆	v ₆	-	-	-	-	-	-	-	sim	sim	sim
	Hiberna	p ₇	v ₇	v ₇	-	-	-	-	-	-	-	não	sim	não
Ações	Classe				*g ₁	peixe	mamífero	réptil	pássaro	anfíbio	?	*g ₁	*g ₂	*g ₃
Funções adaptativas	Anterior	f ₁	B	√	√	√					√			
	Posterior													
	Parâmetros	p ₁	P								p ₁			
				
		p ₇	P								p ₇			
	Variáveis	v ₁	V											
												
	v ₇	V												
Geradores	g ₁	G												

Em trabalhos futuros, deseja-se incluir no modelo uma memória para armazenar as operações na base de regras adaptativas. Essa

memória representa o histórico das operações e as informações contidas na memória, que podem ser utilizadas futuramente para melhorar o modelo de aprendizagem através da experiência. Outros experimentos devem ser realizados para comparar os resultados com modelos de aprendizagem incremental que não utilizam adaptatividade, tais como o C4.5 [12]. Outras funções adaptativas devem ser declaradas e

REFERÊNCIAS

[1] CAVALIN, P.R., SABOURIN, R., SUEN, C.Y., BRITTO JR., A.S.: "Evaluation of Incremental Learning Algorithms for An HMM-Based Handwritten Isolated Digits Recognizer". In: The 11th International Conference on Frontiers in Handwriting Recognition (ICFHR 2008), Montreal, 19-21, 2008.
 [2] DUDA, R. O.; HART, P. E.; STORK, D. G.: *Pattern Classification*. 2ª Edição. Wiley, 2001. ISBN: 0471056693.

- [3] HUGHES, M. L., SHANK, R. M., STEIN, E. S.: *Decision Tables*. Midi Publications, Management Development Institute, Divisions of Information, Industries, Inc., Wayne, Pennsylvania, 1968.
- [4] IWAJ, M. K.: “Um formalismo gramatical adaptativo para linguagens dependentes de contexto”. Tese (Doutorado), EPUSP. São Paulo, 2000.
- [5] NETO, J. J.: “Solving complex problems with Adaptive Automata”. Lecture Notes in Computer Science. S. Yu, A. Paun (Eds.): Implementation and Application of Automata, CIAA 2000, Vol.2088, London, Canada, Springer-Verlag, pp.340, 2000.
- [6] NETO, J. J. “Adaptive Rule-Driven Devices – General Formulation and Case Study”. Revista de Engenharia de Computação e Sistemas Digitais, São Paulo, v.1, n.1, 2001.
- [7] NETO, J. J.: “Contribuição à metodologia de construção de compiladores”. São Paulo, 272p. Tese (Livre-Docência), Escola Politécnica, Universidade de São Paulo, 1993.
- [8] PISTORI, H. “Tecnologia Adaptativa em Engenharia de Computação: estado da arte e aplicações”. Tese (Doutorado), Escola Politécnica da USP, 2003.
- [9] PISTORI, H.; NETO, J. J.: “Decision Tree Induction using Adaptive FSA”. CLEI Electronic Journal. Volume 6, Number 1, 2003
- [10] PRATI, R. C.: “Novas abordagens em aprendizado de máquina para a geração de regras, classes desbalanceadas e ordenação de casos”. Tese, ICMC-USP, 2006.
- [11] QUINLAN, J. R.: “Induction of decision trees”. Machine Learning, 1, 81-106, 1986
- [12] QUINLAN, J. R.: “C4.5: Programs for Machine Learning”. Morgan-Kaufmann, San Francisco, 1993.
- [13] TAN, P. STEINBACH, M.; KUMAR, V.: *Introduction to Data Mining*. Boston, MA, USA Addison-Wesley Longman Publishing Co., Inc., 2005.
- [14] THEODORIDIS, S.; KOUTROUMBAS, K.: *Pattern Recognition*. 3ª Edição. Academic Press, 2006.
- [15] TCHEMRA, A. H.: “Tabela de Decisão Adaptativa na Tomada de Decisão Multicritério”. Tese (Doutorado), EPUSP, São Paulo, 2009.
- [16] RUSSEL, S. J.; NORVIG, P.: *Artificial Intelligence: A Modern Approach*. 2ª Edição. Prentice-Hall, 2002. ISBN - 10: 0137903952

Renata Luiza Stange é mestranda em Engenharia Elétrica (Área de concentração: Computação e Sistemas Digitais) pela Escola Politécnica da Universidade de São Paulo. Especialista em Administração de Sistemas de Informação pela Universidade Federal de Lavras e Bacharel em Análise de Sistemas pela Universidade Estadual do Centro-Oeste do Paraná. Atualmente é professora da Universidade Federal Tecnológica do Paraná.



João José Neto é graduado em Engenharia de Eletricidade (1971), mestre em Engenharia Elétrica (1975), doutor em Engenharia Elétrica (1980) e livre-docente (1993) pela Escola Politécnica da Universidade de São Paulo. Atualmente, é professor associado da Escola Politécnica da Universidade de São Paulo

(EPUSP) e coordena o LTA – Laboratório de Linguagens e Tecnologia Adaptativa do Departamento de Engenharia de Computação e Sistemas Digitais da EPUSP. Tem experiência na área de Ciência da Computação, com ênfase nos Fundamentos da Engenharia da Computação, atuando principalmente nos seguintes temas: dispositivos adaptativos, tecnologia adaptativa, autômatos adaptativos, e em suas aplicações à Engenharia de Computação, particularmente em sistemas de tomada de decisão adaptativa, análise e processamento de linguagens naturais, construção de compiladores, robótica, ensino assistido por computador, modelagem de sistemas inteligentes, processos de aprendizagem automática e inferências baseadas em tecnologia adaptativa.

