

## Classificação de imagens pelo método de Árvore de Decisão Oblíqua

Ivo Paixão de Medeiros<sup>1</sup>  
Carlos Alberto Pires de Castro Filho<sup>2,3</sup>  
Guaraci José Erthal<sup>2</sup>  
Luciano Vieira Dutra<sup>2</sup>

<sup>1</sup> Instituto Tecnológico da Aeronáutica - ITA  
Praça Marechal Eduardo Gomes, 50 – Vila das Acácias - 12228-900  
São José dos Campos - SP, Brasil  
ivopdm@gmail.com

<sup>2</sup> Instituto Nacional de Pesquisas Espaciais - INPE  
Caixa Postal 515 - 12227-010 - São José dos Campos - SP, Brasil  
{pires, gaia, dutra}@dpi.inpe.br

<sup>3</sup> Diretoria de Serviço Geográfico - DSG  
Quartel General do Exército – Bloco “F” – 2º Piso  
Setor Militar Urbano - 70630-910 – Brasília - DF, Brasil

**Abstract.** The Oblique Decision Tree (ODT) classification method has the advantage of dividing feature spaces using multidimensional hyperplanes that are oblique to the Cartesian axes. Because it is a non-parametric classification method, it also is capable of classifying images containing different statistical distribution. This paper aims to present a model of an ODT and perform tests comparing its classification results with other traditional classification methods. The ODT developed is binary and uses the Exchange Method for splitting each node into two subsets of classes, along with Fisher's Linear Discriminant to calculate the parameters for the hyperplanes. Tests have been done using Polarimetric Interferometric and Synthetic Aperture Radar data from the Brazilian Terrestrial Cartography Subproject, also known as “Amazon Radiography”, of the Geographic Service of Brazilian Army (DSG). Throughout the tests the ODT showed slightly better classification results compared to the Ordinary Binary Classification Tree, obtaining better overall accuracy (86.06%) and smaller tree. Besides that, the ODT showed better results than those obtained with traditional classifiers such as the Maximum Likelihood (85.44%), Nearest Neighbor (77.07%) and Mahalanobis Distance (78.72%). In the other hand, the Support Vector Machine classification method obtained higher overall accuracy (87.37%) although a much higher processing time is needed.

**Palavras-chave:** pattern recognition, classification, decision tree, reconhecimento de padrões.

### 1. Introdução

Na disciplina de reconhecimento de padrões, a classificação consiste em atribuir a cada objeto de interesse, um rótulo ou identificador de classe (Theodoridis e Koutroumbas, 2006). Esta associação ocorre em função das características do objeto a ser classificado. No caso de imagens digitais estas características podem ser atributos associados às respostas espectrais, textura, forma e contexto, dentre outros (Pal e Mather, 2001). Métodos supervisionados de classificação pressupõem a existência de objetos rotulados (objetos cuja classe é conhecida) para a construção do classificador. Um desses procedimentos de classificação é o de árvore de decisão.

De acordo com Quinlan (1993), este procedimento tem como vantagem o fato de possuir natureza e propriedades não-paramétricas, podendo classificar imagens com distribuições estatísticas diferentes da gaussiana, heterogêneas e possuidoras de ruídos. Além disso, podem também ser manipulados outros tipos de dados cujos valores são diferentes dos numéricos. Já, segundo Friedl e Brodley (1997), as árvores de decisão têm como vantagem o fato de serem simples e flexíveis, realizando testes sequenciais e de fácil compreensão cuja semântica é praticamente intuitiva.

Uma árvore de decisão é constituída internamente de: (a) nós de decisão, que particionam o espaço de atributos através de superfícies de decisão e (b) nós terminais, ou “folhas”, que permitem associar uma classe a cada objeto (vetor de atributos:  $\mathbf{x}$ ). Para atributos numéricos, utilizam-se comumente funções lineares como modelos para as superfícies de decisão. Serão retas para espaços de dois atributos, planos no caso 3-D, e em geral (para  $d$  atributos) serão hiperplanos.

A Figura 1 ilustra o uma árvore de decisão univariada (ou OBCT - *Ordinary Binary Classification Tree*), suas regras de classificação e a respectiva representação da classificação dos dados em um espaço de atributos. Este é o caso mais simples de árvore de decisão, sendo chamada de univariada em função de suas regras estarem associadas somente a um único atributo.

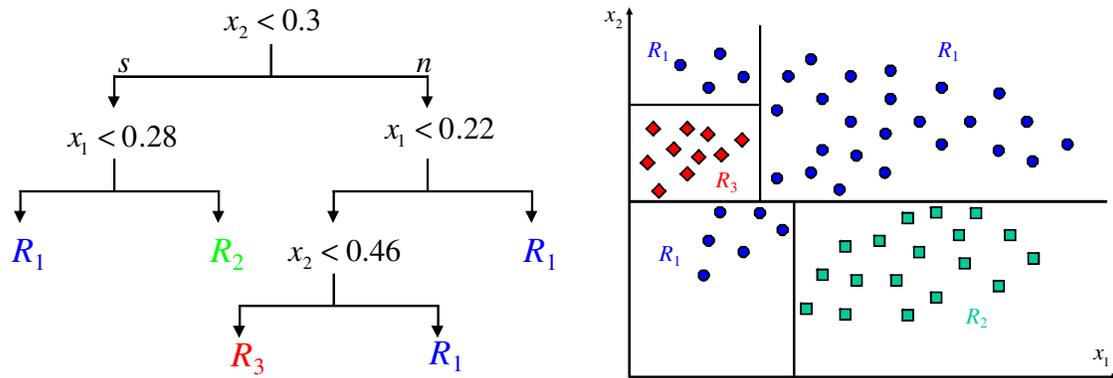


Figura 1. Exemplo de árvore de decisão univariada. Adaptado de Duda et al. (2001)

Em cada nó da OBCT, o hiperplano de decisão intercepta apenas um dos eixos coordenados, e seu modelo, para um dado atributo  $x_j$  é descrito pela Equação 1 onde  $\mathbf{x}$  é o vetor de atributos do objeto, e o hiperplano intercepta o eixo  $j$  em  $-w_0$ .

$$f(\mathbf{x} | j, w_0) = x_j + w_0 = 0 \quad (1)$$

Já, na Figura 2 a árvore de decisão é chamada de oblíqua (ODT - *Oblique Decision Tree*) ou multivariada linear (Yildiz e Alpaydin, 2005), porque os testes envolvem hiperplanos com orientação oblíqua aos eixos dos atributos. No caso da ODT, o modelo adotado para as superfícies de decisão é dado pela Equação 2 onde  $m$  é o nó de decisão,  $\mathbf{w}_m$  é o vetor de pesos do hiperplano ( $d$ -dimensional),  $w_{m0}$  o valor limitante quando  $\mathbf{x}$  está projetado sobre  $\mathbf{w}_m$  e  $T$  indica o vetor transposto.

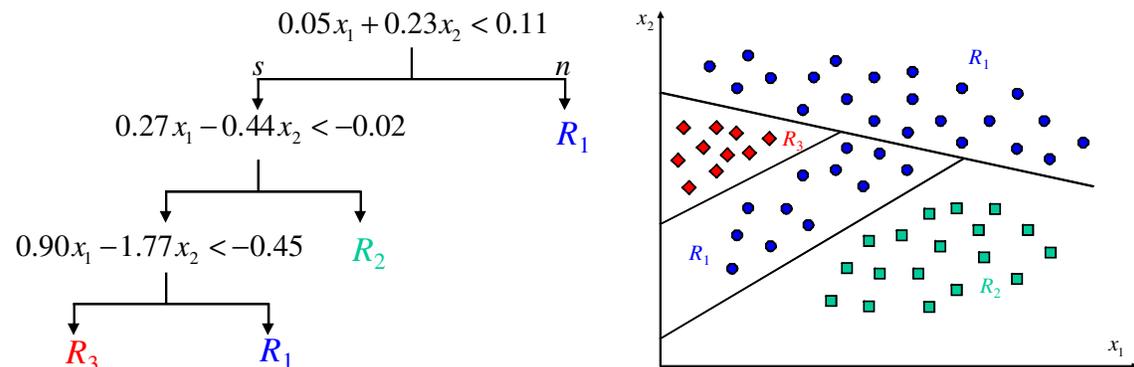


Figura 2. Exemplo de árvore de decisão oblíqua. Adaptado de Duda et al. (2001)

$$f_m(\mathbf{x} | \mathbf{w}_m, w_{m0}) = \mathbf{w}_m^T \mathbf{x} + w_{m0} \quad (2)$$

Tanto na OBCT como na ODT o teste a ser realizado em cada nó de decisão pode ser descrito como uma função booleana da forma: se  $f_m(\mathbf{x} | \Phi) > 0$  então siga o galho da direita; caso contrário, siga o galho de esquerda (Yildiz e Alpaydin, 2005). A classificação de um padrão  $\mathbf{x}$  consiste em apresentá-lo ao primeiro nó de decisão da árvore (chamado de “raiz”), realizar o teste, seguir para o nó do galho selecionado, e sucessivamente aplicar testes até atingir um nó folha onde sua classe será definida.

Este trabalho tem por objetivos apresentar uma modelagem de uma árvore de decisão oblíqua. Além disto, serão realizados testes e feitas análises visando comparar os resultados com outros métodos de classificação, dentre os quais uma árvore de decisão univariada.

## 2. O Programa

O programa, desenvolvido na linguagem de programação IDL (*Interactive Data Language*), permite a entrada de imagens multi-canais, admitindo resolução radiométrica de 8 bits por píxel na versão atual. Logo após a seleção da imagem, é necessário que o usuário realize a coleta das amostras de treinamento, visto que o método de classificação a ser utilizado é supervisionado.

O algoritmo principal foi decomposto em módulos, sendo os principais: módulo de construção da árvore de decisão; módulo de construção de nós de decisão; módulo de cálculo do hiperplano; e módulo de classificação.

No módulo de construção da árvore de decisão inicialmente são gerados números identificadores para os nós que são criados a partir de cada “nó mãe”, ou nó de origem. Os novos nós são colocados então em uma fila e a futura divisão de cada nó desta fila somente ocorrerá caso sejam atendidos três critérios, chamados critérios de pré-poda: 1) caso o nó não seja puro, isto é, somente poderá ser dividido se naquele nó houver instâncias rotuladas a duas ou mais classes distintas; 2) caso a porcentagem mínima de instâncias estiver dentro da tolerância definida pelo usuário; e 3) caso a porcentagem mínima de diminuição de entropia estiver dentro do valor definido pelo usuário.

O primeiro passo do módulo de construção de nós de decisão consiste na definição do nó raiz da árvore. Com base no conjunto completo de treinamento, e supondo-se que existem apenas duas classes, seleciona-se o hiperplano discriminante que satisfaz o critério de Fisher. Deste modo os dados serão divididos em dois conjuntos sendo cada um associado a um dos nós filhos. Como é comum o número de classes  $k > 2$ , é preciso transformar o problema de  $k$  classes em um problema de duas classes. Para isto é utilizado o Método de Troca (Yildiz e Alpaydin, 2005), onde inicialmente são formados dois conjuntos de classes “sementes”, cujas médias das classes encontram-se mais distantes entre si. Inicia-se então o processo de troca onde cada uma das classes é retirada do seu conjunto original e alocada no outro. Calcula-se então o discriminante de Fisher para esta configuração de classes. Em seguida, calcula-se a variação da entropia calculada para os dados antes e depois da divisão. A classe que, em caso de troca, ocasionar a maior diminuição no valor de entropia é então efetivamente trocada, passando a fazer parte do outro conjunto de classes. Ao término do processo, quando nenhuma troca de classes entre os dois conjuntos diminui o valor de entropia, o hiperplano encontrado é associado ao nó corrente. Este módulo é ativado enquanto houver nós a serem particionados.

O módulo de cálculo do hiperplano consiste na determinação dos parâmetros  $\mathbf{w}$  e  $w_0$ , a partir de dois conjuntos de dados. Vamos chamá-los de *esquerda* ( $C_E$ ) e *direita* ( $C_D$ ). A idéia é separar os dois conjuntos por um hiperplano que satisfaz o critério de Fisher. O critério de Fisher permite encontrar uma direção  $\mathbf{w}$  que maximiza a separabilidade entre duas classes

sobre ela projetadas. Esta direção define a orientação de um hiperplano, e sua posição é obtida pelo parâmetro  $w_0$  ser for considerada a hipótese Gaussiana para os dados. Para tanto os seguintes parâmetros são calculados: vetores de médias ( $\mathbf{m}_E$  e  $\mathbf{m}_D$ ) e matrizes de espalhamento ( $S_E$ ,  $S_D$ ) para os 2 grupos; e matriz de espalhamento total ( $S = S_E + S_D$ ). Os tamanhos dos conjuntos são dados por  $n_E$  e  $n_D$ . A solução de Fisher pode então ser calculada através das Equações 3 e 4 e a função discriminante será dada pela Equação 2 apresentada anteriormente.

$$\mathbf{w} = S^{-1}(\mathbf{m}_E - \mathbf{m}_D) \quad (3)$$

$$w_0 = -\frac{1}{2}(\mathbf{m}_E + \mathbf{m}_D)^T S^{-1}(\mathbf{m}_E + \mathbf{m}_D) + \log \frac{n_E}{n_D} \quad (4)$$

No último módulo, a classificação da imagem é realizada pixel a pixel. Cada pixel, sendo representado pelo seu vetor de atributos ( $\mathbf{x}$ ), é submetido às regras lógicas do tipo se-então-senão de cada nó. Desta forma cada instância percorre um caminho na árvore até atingir uma folha correspondente à uma determinada classe a qual aquele pixel ficará associado.

A complexidade do algoritmo desenvolvido depende de três variáveis:  $N$  (número total de pixels na imagem),  $k$  (número de classes definidas na etapa de treinamento) e  $n$  (número total de pixels na amostra de treinamento). Somando a complexidade das etapas de construção das regras de classificação e a de classificação da imagem o algoritmo possui final complexidade de  $O(n.k^2 + N.\log_2^N)$ .

### 3 Testes Comparativos com Outros Tipos de Classificadores

Inicialmente foi realizado um teste comparativo entre árvores univariada (OBTD) e oblíqua (ODT). Neste teste optou-se por utilizar dados PolInSAR – Radar de Abertura Sintética Polarimétrico e Interferométrico - do Projeto Cartografia da Amazônia, mais especificamente o Subprojeto Cartografia Terrestre em execução pela Diretoria de Serviço Geográfico (DSG) do Exército Brasileiro. A área de trabalho situa-se próxima a cidade de Barcelos – AM, com coordenadas centrais de 63°15'11''w/ 0°44'46''s, no sistema geodésico WGS-84, e datadas de 30 de março de 2009.

Após a extração de diversos atributos dos dados PolInSAR, buscou-se utilizar somente dois atributos para facilitar a análise visual da divisão do espaço de atributos. Também visando esta análise visual no espaço bidimensional, foram definidas somente quatro classes de uso do solo: água, campinarana, floresta primária alagada e floresta primária de solo firme. A quantidade de amostras de treinamento para cada classe foi de 100 pixels. Para maiores detalhes sobre o processo de extração dos atributos SAR e do processo de definição das classes de trabalho, sugere-se consultar Castro Filho e Santos (2010).

Os dois atributos definidores do espaço a ser analisado foram selecionados através da utilização de técnicas de seleção de atributos no programa Weka (*Waikato Environment for Knowledge*). Neste caso o método de busca utilizado foi o de ranqueamento e os de avaliação foram os de *Information Gain*, *Chi Square* e *One R*. Em todos os casos os atributos melhores avaliados foram os de P\_Coh (coerência interferométrica da banda P) e de Hint (altura interferométrica). Para maiores detalhes sobre métodos de seleção de atributos sugere-se consultar Theodoridis e Koutroumbas (2006).

O algoritmo de classificação pelo método de árvore de decisão univariada utilizado foi o J4.8, que é o nome dado à implementação da tradicional árvore C4.5 no Weka. O parâmetro de pré-poda desta árvore foi de duas instâncias como tamanho mínimo para compor cada nó. A árvore de decisão univariada construída e a respectiva divisão do espaço de atributos encontram-se ilustradas nas Figuras 3 e 4, respectivamente. Nela foram utilizadas nove regras de decisão para dividir o espaço de atributos também em nove sub-espacos.

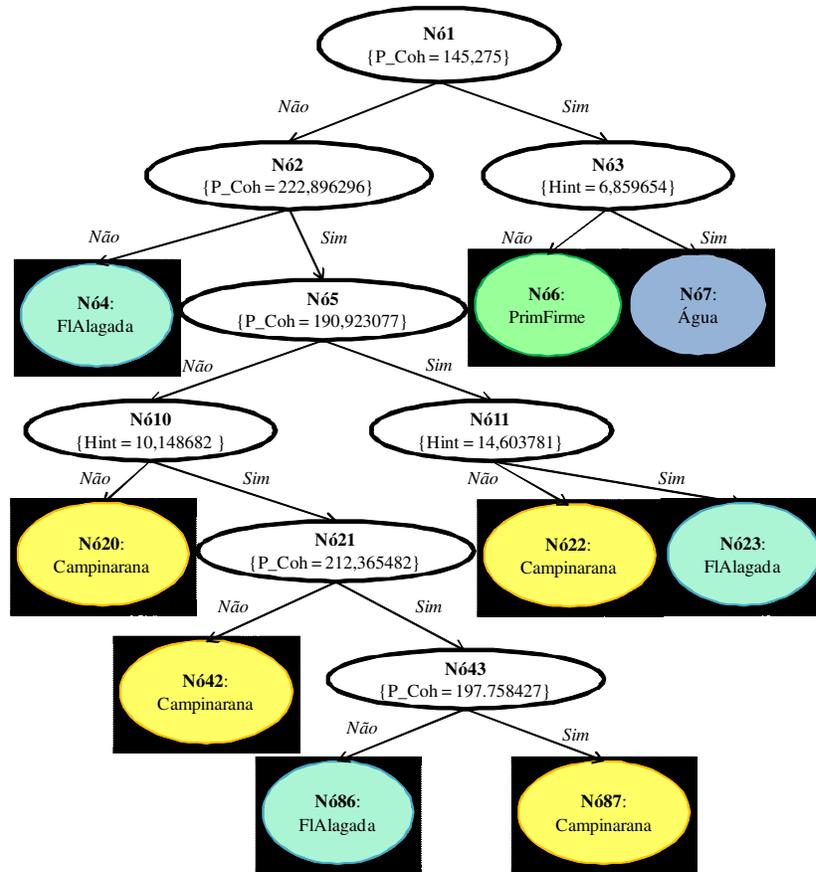


Figura 3. Árvore de Decisão Univariada gerada no Weka.

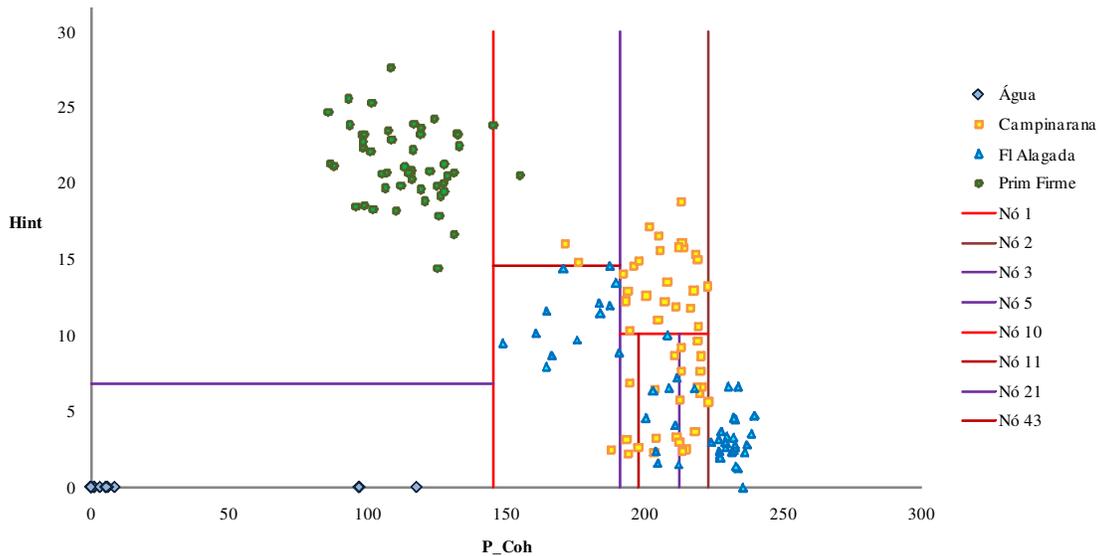


Figura 4. Divisão do Espaço de Atributos pela Árvore de Decisão Univariada.

Já, para a árvore de decisão oblíqua, o parâmetro de pré-poda foi equivalente ao utilizado na árvore univariada: 2% de tamanho mínimo de instâncias para compor cada nó. Neste caso o parâmetro de porcentagem mínima de diminuição de entropia foi desconsiderado. A árvore construída e a divisão do espaço de atributos encontram-se ilustradas, respectivamente, nas Figuras 5 e 6. Esta árvore oblíqua possui cinco regras de decisão para definir seis sub-espacos.

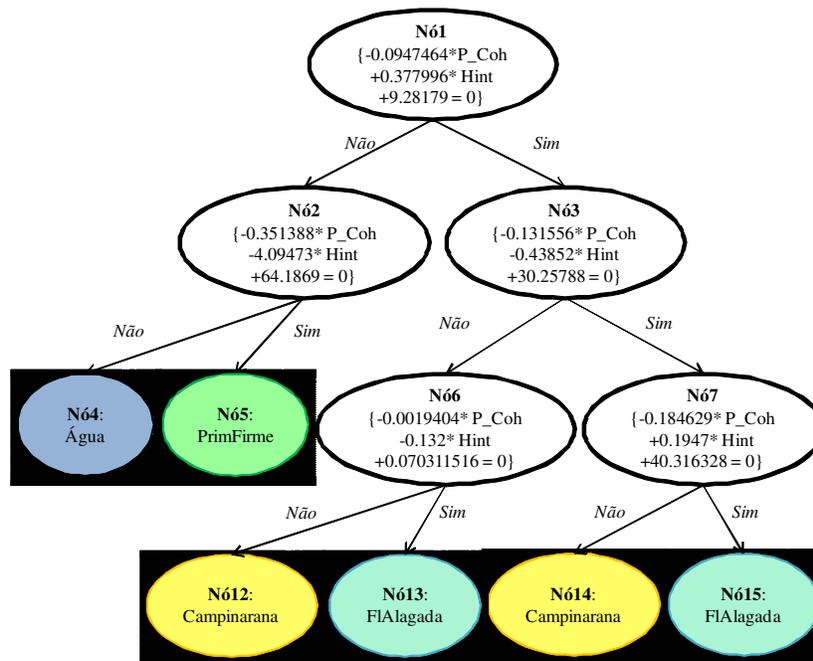


Figura 5. Árvore de Decisão Oblíqua.

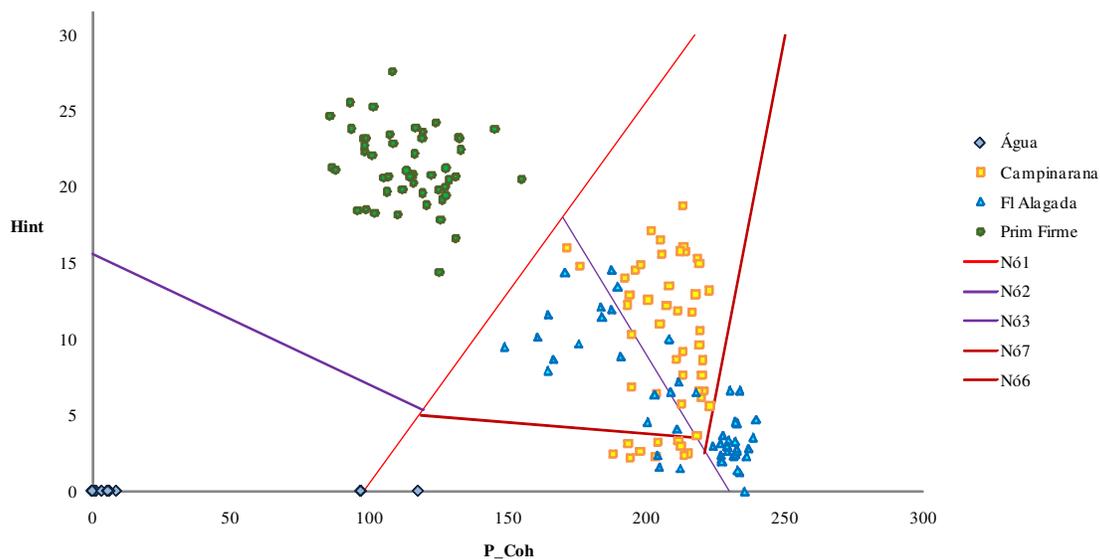


Figura 6. Divisão do Espaço de Atributos pela Árvore de Decisão Oblíqua.

As etapas de classificação e a respectiva avaliação foram realizadas no sistema de processamento de imagens ENVI (*ENvironment for Visualizing Images*) versão 4.7. Os resultados das classificações encontram-se na Figura 7. Qualitativamente, através de uma análise visual, observa-se que a classificação através da árvore de decisão univariada apresentou maior quantidade de píxeis isolados em todas as classes. Isto se deve ao fato de que a árvore construída (Figura 3) possui um número maior de nós de decisão do que a árvore desenvolvida pelo método oblíquo (Figura 5), gerando regras mais específicas.

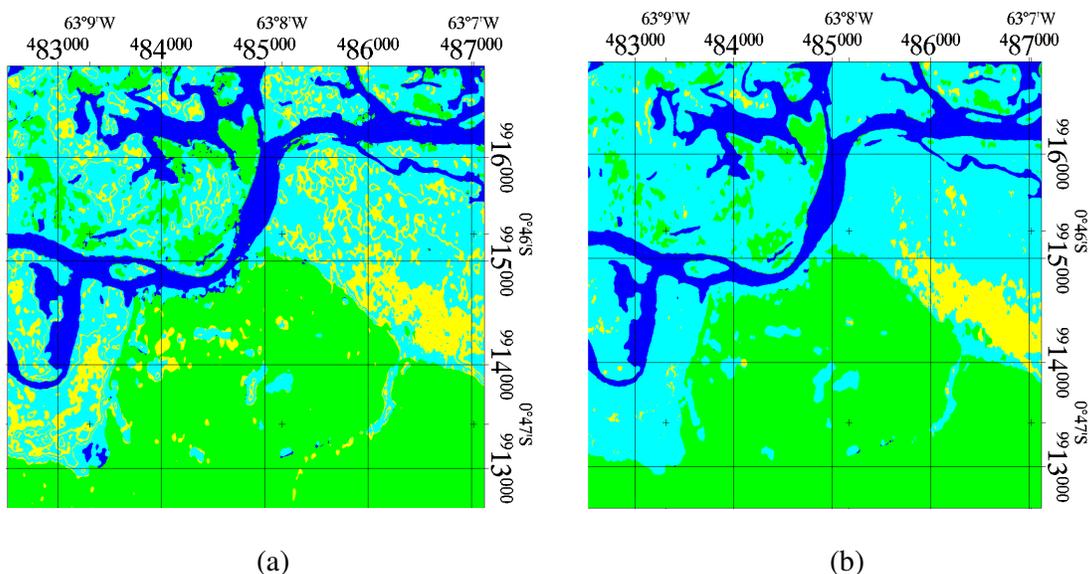


Figura 7. Imagem classificada através de árvore de decisão univariada (a) e através de árvore de decisão oblíqua (b) com coordenadas em UTM/WGS-84.

A avaliação quantitativa, para a etapa de teste, foi realizada através da coleta de 10.000 píxeis por classe e definindo-os como referência. A matriz de confusão com valores percentuais desta avaliação encontram-se na Figura 8. Em ambas as matrizes observa-se que as classes de água e de floresta primária firme (PrFm) foram classificadas com porcentagens próximas a 100%. Por outro lado houve confusão entre as classes de campinarana (Camp.) e de floresta alagada (FIAI.). A acurácia global e índice Kappa da classificação através de árvore de decisão oblíqua foram próximos, porém superiores, aos obtidos pela árvore de decisão univariada.

Classif\Referência	Classif\Referência				Classif\Referência				
	Água	Camp.	FIAI.	PrFm.	Água	Camp.	FIAI.	PrFm.	
Água	100.00	0.00	0.00	0.00	Água	100.00	0.00	0.00	
Campinarana	0.00	65.69	20.35	4.08	Campinarana	0.00	64.35	13.45	
FIAIagada	0.00	34.31	78.78	0.00	FIAIagada	0.00	35.65	84.04	
PrimFirme	0.00	0.00	0.87	95.92	PrimFirme	0.00	0.00	2.51	
Acurácia Global = <b>84.03%</b> Índice Kappa = 0.7806					Acurácia Global = <b>86.60%</b> Índice Kappa = 0.8149				
(a)					(b)				

Figura 8. Matrizes de confusão de teste da árvore de decisão univariada (a) e da árvore de decisão oblíqua (b).

Além deste teste comparativo entre diferentes árvores de decisão, foram realizadas também classificações utilizando os seguintes métodos: MaxVer (Máxima Verossimilhança), Vizinho Mais Próximo (*Nearest Neighbor*, 1-NN), Distância de Mahalanobis e SVM (*Support Vector Machine*). Para a classificação pelo método SVM foi adotado o método linear com parâmetro de penalidade  $C = 100$ . Estes métodos foram selecionados e aplicados em função da disponibilização de seus algoritmos no ENVI 4.7.

A quantidade de amostras de treinamento, de teste e as avaliações quantitativas foram realizadas igualmente às descritas anteriormente no teste comparativo entre OBDT e ODT. As acurarias globais e índices Kappa (entre parênteses) obtidos foram os seguintes: 85.44% (0.7949) para a classificação por MaxVer; 77.07% (0.6882) para o Vizinho Mais Próximo;

78.72% (0.7082) para a Distância de Mahalanobis; e 87.37% (0.8255) para o SVM. Logo, sobre a imagem de SAR o melhor resultado foi obtido através do classificador SVM, seguido da Árvore de Decisão Oblíqua. Neste caso, além de ter sido observada a presença de ruído no dado a ser classificado (ruído *speckle* inerente à natureza do dado SAR), o mesmo possui distribuição estatística que varia em função de diversos fatores, possível motivo pelo qual o resultado no método MaxVer, implementado com base em uma distribuição gaussiana, ter sido inferior.

#### 4 Conclusões

Na comparação entre a árvore de decisão oblíqua e a univariada, a árvore oblíqua apresentou uma quantidade menor de nós, dividindo o espaço de atributos em uma menor quantidade de sub-espacos e gerando uma regra de classificação mais simples. A acurácia global e índice Kappa atingidos em ambos os métodos foram equivalentes, com uma pequena melhora pela árvore de decisão oblíqua.

Apesar das vantagens, observa-se que os cálculos matemáticos para cada nó de decisão nas árvores oblíquas são mais complexos do que os da univariadas por envolverem todos os atributos em trabalho. Além deste custo computacional maior em cada nó, o usuário perde também a intuitividade ao analisar cada regra.

Comparando com outros tipos de classificadores tradicionais é possível observar que o método de classificação por árvore de decisão, por ser não-paramétrico, obteve destaque sobre a classificação da imagem SAR cuja distribuição estatística é diferente da gaussiana. Apesar de não ter obtido resultados superiores ao do SVM, o tempo de processamento foi inferior ao daquele método. Este baixo custo computacional é evidenciado pela simplicidade na complexidade do algoritmo desenvolvido.

#### Referências

Castro Filho, C. A. P.; Santos, J. R. Classificação de imagens PolInSAR utilizando técnicas de mineração de dados. **Ambiência – Revista do Setor de Ciências Agrárias e Ambientais**. v. 6, Especial. 2010. 33-43p. ISSN 1808-0251.

Duda, R.O.; Hart, P.E.; Stork, D.G. **Pattern Classification**. 2nd ed. John Wiley & Sons. NY. 2001. 513p.

Friedl, M. A.; Brodley, C. E. Decision tree classification of land cover from remotely sensed data. **Remote Sensing of Environment**. v. 61. 1997. Pg 399-409.

Pal, M.; Mather, P. M. Decision Tree Based Classification of Remotely Sensed Data. In: 22<sup>a</sup> Asian Conference on Remote Sensing, 2001, Singapore. **Anais...** Disponível em: <<http://www.crisp.nus.edu.sg/~acrs2001/pdf/046PAL.PDF>> . Acesso em 02 Out 2009.

Quinlan, J. R. **C4.5: Programs for Machine Learning**. Morgan Kaufmann. California. 1993. 235p.

Theodoridis, S.; Koutroumbas, K. **Pattern Recognition**. 3. ed. San Diego: Academic Press, 2006. 651 p.

Yildiz, O. T.; Alpaydin, E. Linear discriminant trees. **International Journal of Pattern Recognition and Artificial Intelligence**. v. 19, n. 3. 2005. 763-782p.