

# Seleção de Atributos Utilizando Busca Exaustiva e Máquina de Vetores Suporte

Eliana Pantaleão<sup>1</sup>, Luciano V. Dutra<sup>1</sup>, Sérgio L. Dousseau<sup>2</sup>

<sup>1</sup>Divisão de Processamento de Imagens – Instituto Nacional de Pesquisas Espaciais (INPE)

São José dos Campos – SP – Brazil

<sup>2</sup>Fundação de Ciência Aplicações e Tecnologia Espaciais (FUNCATE)

São José dos Campos – SP – Brazil

<sup>1</sup>{elianap, dutra}@dpi.inpe.br, <sup>2</sup>sergio@funcate.org.br

**Abstract.** *This paper describes the use of a Support Vector Machine (SVM) classifier to select attributes by exhaustive search. The module, implemented in IDL, was added to an image processing system developed by INPE students.*

**Resumo.** *Este artigo descreve a aplicação de um classificador por máquina de vetores suporte, ou SVM (Support Vector Machine), na seleção de atributos através de busca exaustiva. O módulo, implementado em IDL, foi incorporado a um sistema de processamento de imagens já existente, desenvolvido por alunos do INPE.*

## 1. Introdução

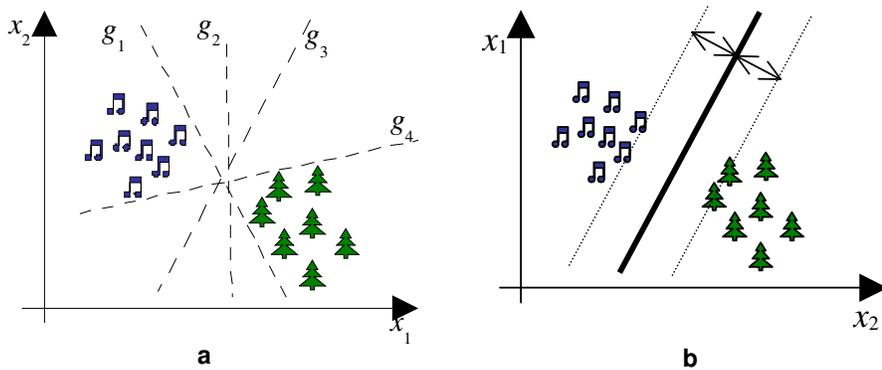
No processamento de imagens digitais é comum o uso do processo de extração de atributos, com o intuito de descrever informações sobre determinadas características da imagem. Em sensoriamento remoto, são frequentemente utilizados atributos para descrever texturas, componentes principais e imagens-fração, por exemplo. Se o aumento do número de atributos fornece mais informações sobre os padrões que se deseja classificar, também acrescenta informações redundantes e ruidosas, que podem degradar o processo de classificação, conforme descrito pelo fenômeno denominado “maldição da dimensionalidade”. Além disso, o tempo computacional utilizado na classificação também aumenta com o número de atributos utilizado.

## 2. O Classificador SVM

Uma máquina de vetores suporte, ou SVM (Support Vector Machine) é uma máquina de aprendizagem universal cuja superfície de decisão é parametrizada por um conjunto de vetores suporte, e por um conjunto de pesos correspondentes. SVM's podem ser utilizadas no reconhecimento de padrões, e também para aproximar funções reais, estimar regressões e no processamento de sinais (Vapnik, 1996).

No reconhecimento de padrões, o objetivo do classificador é encontrar um hiperplano que classifica corretamente todos os vetores de treinamento. Este hiperplano não é único, como mostra a Figura 1a, que ilustra duas classes linearmente separáveis e quatro possíveis separadores lineares. Qualquer um dos quatro hiperplanos separadores

mostrados na Figura 1a pode ser utilizado para separar as classes nos dados de treinamento. Entretanto, quando o classificador é utilizado na prática, os dados utilizados não pertencem ao conjunto de treinamento. Embora haja grande probabilidade de que se localizem próximos aos dados existentes, pequenas variações são esperadas. As máquinas de vetores suporte fornecem o hiperplano separador ótimo entre duas classes, ou seja, a superfície que deixa a maior margem entre as duas classes, como mostra a Figura 1b. Assim, os dados possuem mais “espaço” para se localizarem sem que haja erros de classificação (Webb, 2002). Os vetores de treinamento atravessados pelas linhas pontilhadas na Figura 1b são os vetores suporte.



**Figura 1: duas classes linearmente separáveis, quatro possíveis separadores lineares e o hiperplano separador ótimo**

A regra de decisão pode ser expressa por: (Theodoridis, 2003)

$$\mathbf{w}^T \mathbf{x} + w_0 \begin{cases} > 0 \\ < 0 \end{cases} \Rightarrow \mathbf{x} \in \begin{cases} \omega_1, \text{ com valor numérico correspondente } y = +1 \\ \omega_2, \text{ com valor numérico correspondente } y = -1 \end{cases}$$

Em muitos casos reais não há um limite linear separando as classes, logo, não faz sentido procurar um hiperplano separador ótimo. Uma forma de lidar com dados não linearmente separáveis é através da relaxação das restrições, através da introdução de variáveis de folga,  $\xi_i$ , e de um valor de penalidade em caso de classificação incorreta,  $C$ . A função de custo passa a incluir um fator de regularização,  $C \sum_i \xi_i$ , que pode ser interpretado como uma medida da classificação incorreta. Quanto menor o valor de  $C$ , menor é a penalidade aplicada aos *outliers*.

O caso não linear também pode ser abordado através do uso de mapeamentos em um espaço de atributos de maior dimensionalidade, no qual as classes sejam linearmente separáveis. Em geral, são utilizadas, para este fim, funções simétricas que satisfaçam as condições de Mercer (Theodoridis, 2003), comumente denominadas funções de kernel.

As máquinas de vetores suporte foram projetadas para separar apenas duas classes. Para um número maior de classes, é necessário utilizar uma estratégia de integração dos hiperplanos gerados. Na estratégia um-contra-todos, para  $C$  classes, são construídos  $C$  classificadores binários, cada um dos quais é treinado para diferenciar uma classe de todas as demais, gerando uma imagem com valores discriminantes. A classificação utiliza o maior valor discriminante para designar a classe do padrão.

### **3. Seleção de Atributos e Busca Exaustiva**

O critério ótimo para realizar a seleção de atributos é minimizar o erro de classificação. O conjunto ideal de atributos é aquele que, quando utilizado para classificar a imagem, fornecerá a menor quantidade de pixels classificados incorretamente. Entretanto, a solução sub-ótima pode não apresentar diferença significativa em relação à ótima, e utilizar menor número de atributos, diminuindo o tempo computacional gasto na classificação. A busca exaustiva gera todas as combinações possíveis de atributos e calcula o erro de classificação para cada conjunto. A seleção do resultado utiliza um teste de hipótese para a significância da melhoria obtida ao ser adicionado um atributo.

### **4. Implementação, Resultados e Conclusões**

O método de classificação utilizando máquinas de vetores suporte foi implementado na linguagem IDL e incorporado ao sistema SCID (Sistema de Classificação de Imagens Digitais), desenvolvido por alunos de pós-graduação do INPE.

Os dados de entrada são: a imagem original, com P canais, o conjunto de treinamento, com n amostras e os parâmetros do classificador: o valor limitante C para os multiplicadores de Lagrange, o tipo de kernel a ser utilizado e os parâmetros do mesmo. Para resolver o problema de otimização, foi utilizado o recurso IDL denominado CONSTRAINED\_MIN, um procedimento que resolve problemas de otimização não lineares, tanto de minimização, como de maximização.

Um teste foi realizado com imagens CBERS e dados da Funcate e do PRODES. Foram fornecidos ao sistema seis atributos: as bandas originais 2, 3 e 4 da imagem e as três imagens-fração (solo, vegetação e sombra). O objetivo é a identificação de áreas de rebrota, na área dos municípios de Nortelândia e Arenópolis (MT).

Os resultados mostraram que a melhor opção é utilizar apenas a imagem-fração vegetação, pois, embora resulte em um erro de classificação maior que outros conjuntos, fornece a melhor relação custo/benefício por utilizar apenas um atributo. Este resultado coincide com a análise dos dados realizada anteriormente, e o procedimento pode ser estendido para outras regiões de floresta.

### **Referências**

- Burges, C. J. C. Simplified support vector decision rules. In L. Saita, editor, Machine Learning. Proceedings of the Thirteenth International Conference (ICML '96), Bari, Italy July 3-6, 1996, p. 71-77, San Francisco, CA, 1996. Morgan Kaufmann.
- Instituto Nacional de Pesquisas Espaciais (INPE). PRODES Digital. São José dos Campos, 2007. <<http://www.obt.inpe.br/prodesdigital/>>. Acesso: 4 set. 2007.
- Theodoridis, S. & Koutroumbas, K. Pattern Recognition. Elsevier, 2003.
- Vapnik, V. N.; Golowich, Steven E. and Smola, Alex. Support vector method for function approximation, regression estimation and signal processing. <http://citeseer.ist.psu.edu/vapnik96support.html>
- Webb, A. Statistical pattern recognition. Wiley, 2002.