

Contexto Espacial na Ponderação de Amostras Semi-Rotuladas

Vitor Haertel ¹

Denis Altieri de Oliveira Moraes ²

¹ Universidade Federal do Rio Grande do Sul - UFRGS
Centro Estadual de Pesquisa em Sensoriamento Remoto e Meteorologia - CEPSRM
Caixa Postal 15052 - 91501-970 - Porto Alegre - RS, Brasil
victor.haertel@ufrgs.br

² Universidade Federal de Santa Maria - UFSM
Centro de Ciências Exatas e Naturais – Departamento de Estatística
Av. Roraima 1000, 97105-900 – Cidade Universitária – Santa Maria – RS, Brasil
daltieri@smail.ufsm.br

Abstract: In this paper it is proposed a new approach to select and weigh semi-labeled samples to be added to the available labeled samples into the process of parameter estimation in parametric classifiers. In the literature it is reported an iterative procedure to deal with semi-labeled samples in an iterative way, making use of the available spectral data. In this study we propose the inclusion of contextual information to further improve the process of selecting semi-labeled samples. The proposed methodology is tested by applying it to hyperspectral image data covering a test area. Initial results are presented and discussed.

Palavras-chave: semi-labeled samples, spatial context, hypespectral image data.

1. Introdução

Neste trabalho é proposta uma nova metodologia para fins de obtenção de amostras semi-rotuladas no processo de estimação de parâmetros em classificadores paramétricos. São também apresentados alguns resultados iniciais obtidos pela aplicação da metodologia proposta a dados em alta dimensionalidade (imagens hiperespectrais).

Classes espectralmente muito semelhantes, isto é, classes que compartilham dos mesmos valores para o vetor de médias podem não obstante ser separadas com um alto grau de acurácia, empregando dados em alta dimensionalidade (hiper-espacos), na condição de que suas matrizes de covariância difiram significativamente (Fukunaga, 1990, Landgrebe, 2003). O emprego de dados em alta dimensionalidade em classificadores paramétricos apresenta, entretanto, dificuldades com relação à estimação dos parâmetros (vetor de médias e matriz de covariância) do classificador. Na medida em que a dimensionalidade dos dados (número de bandas espectrais) cresce, aumenta também o número de parâmetros a serem estimados, especialmente na matriz de covariância. Em situações reais, entretanto, o número de amostras disponíveis para estimação destes parâmetros (amostras rotuladas) é limitado, impedindo frequentemente a obtenção de estimativas confiáveis. Este fato traduz-se no conhecido efeito de Hughes. Inicialmente, o acréscimo de novas bandas espectrais resulta em um gradual incremento na acurácia da imagem temática produzida pelo classificador. Em um determinado ponto o valor da acurácia atinge um máximo para em seguida passar a diminuir, na medida em que novas bandas espectrais continuam sendo adicionadas. Este fenômeno é causado pela gradativa deterioração na confiabilidade dos valores estimados para um número crescente de parâmetros a partir de um número limitado de amostras de treinamento.

Métodos para reduzir os efeitos de fenômeno de Hughes e assim possibilitar um uso mais amplo de imagens hiperespectrais, vêm sendo o objeto de estudo por vários pesquisadores. De um modo geral, as metodologias propostas se inserem em três grandes grupos: (1) análise discriminante regularizada, (2) redução na dimensionalidade dos dados na condição de minimização de perda de informação (seleção/extração de variáveis) e (3) uso de amostras semi-rotuladas. Este trabalho insere-se neste último grupo.

Em uma imagem digital, amostras semi-rotuladas são constituídas por pixels que ocorrem na imagem e que não fazem parte do conjunto das amostras rotuladas (verdade terrestre). O grau de pertinência destes pixels a cada uma das classes em consideração pode ser estimado utilizando abordagens adequadas. Um certo número destas amostras é então adicionado, de uma forma ponderada, ao conjunto das amostras rotuladas. Desta forma, o número total de amostras disponíveis é incrementado, permitindo assim obter estimativas mais confiáveis para os parâmetros do classificador e conseqüentemente imagens temáticas mais acuradas. Na literatura encontram-se relatados estudos desenvolvidos por alguns pesquisadores (Shashahani e Landgrebe, 1994, Jackson e Landgrebe, 2001).

Neste estudo parte-se da abordagem geral proposta por Jackson e Landgrebe (2001), adicionando-se informações de contexto espacial no processo de estimação dos graus de pertinência das amostras semi-rotuladas. A metodologia proposta encontra-se descrita na seção 2 e os testes iniciais objetivando validar a metodologia proposta são apresentados na seção 3.

2. Metodologia

A metodologia proposta encontra-se descrita em Jackson e Landgrebe (2001), Liczbinski e Haertel (2008). Em um passo inicial, os parâmetros do classificador são estimados a partir das amostras rotuladas disponíveis, produzindo uma imagem temática inicial. Amostras semi-rotuladas podem então ser extraídas e seus graus de pertinência estimados por meio da própria função de decisão empregada pelo classificador. Em um processo iterativo os graus de pertinência das amostras semi-rotuladas são atualizados e os parâmetros do classificador são re-estimados, produzindo uma imagem temática mais acurada. Este processo iterativo continua até que um critério de convergência pré-estabelecido seja atingido. Um passo adicional neste processo iterativo foi proposto por Liczbinski e Haertel (2008) para estimação das probabilidades *a priori* das classes empregando as mesmas amostras semi-rotuladas, resultando em um incremento adicional na acurácia da imagem temática produzida pelo classificador.

Neste estudo é proposta uma etapa adicional na estimação dos graus de pertinência das amostras semi-rotuladas. Na metodologia proposta por Jackson e Landgrebe (2001), a ponderação de amostras semi-rotuladas é efetivada empregando informações espectrais somente. Na suposição de normalidade multivariada dos dados, a função de decisão tem a forma descrita na equação (1).

$$G_i(X) = (2\pi)^{-p/2} |\Sigma_i|^{-1/2} \exp[-\frac{1}{2}(X - \mu_i)' \Sigma_i^{-1} (X - \mu_i)] P(\omega_i) \quad (1)$$

Ou, em sua forma logarítmica (computacionalmente mais adequada), conforme a equação (2):

$$G_i(X) = -(X - \mu_i)' \Sigma_i^{-1} (X - \mu_i) - \ln |\Sigma_i| + 2P(\omega_i) \quad (2)$$

Onde μ e Σ representam os dois parâmetros (vetor de médias e matriz de covariâncias) associados à classe ω_i , $P(\omega_i)$ a probabilidade *a priori* e X o padrão (pixel) a ser classificado.

Neste estudo, os graus de pertinência de cada pixel a cada uma das classes são estimados pela correspondente probabilidade *a posteriori* $p(\omega_i / X)$, gerando amostras semi-rotuladas. A probabilidade *a posteriori* pode ser estimada pelo teorema de Bayes, descrito na equação (3):

$$p(\omega_i / X) = \frac{p(X / \omega_i) P(\omega_i)}{\sum_{j=1}^m p(X / \omega_j) P(\omega_j)} \quad (3)$$

Onde m representa o número de classes em consideração.

A probabilidade *a posteriori* (3) pode ser estimada utilizando a função de decisão em sua forma logarítmica (2), resultando na forma descrita na equação (4):

$$w_{i,j} = p(\omega_j / X_i) = \frac{\exp[G_j(X_i)]}{\sum_{k=1}^m \exp[G_k(X_i)]} \quad (4)$$

Onde $w_{i,j}$ representa o grau de pertinência do pixel X_i à classe ω_j .

Para cada uma das m classes, um sub-conjunto das N amostras semi-rotuladas, aquelas com o grau de pertinência mais alto, é adicionado ao conjunto das amostras rotuladas disponíveis, resultando em um número maior de amostras disponíveis para fins de estimação dos parâmetros. Neste caso, no processo de estimação dos vetores de médias (M) e matrizes de covariância (S), as amostras rotuladas entram com peso unitário enquanto as amostras semi-rotuladas entram com um peso igual ao respectivo grau de pertinência. Dessa forma, as equações (5) e (6) definem, respectivamente, o vetor de médias e a matriz de covariâncias das classes:

$$M_j = \frac{1}{\sum_{k=1}^{N_j} w_{k,j}} \sum_{k=1}^{N_j} w_{k,j} X_k \quad (5)$$

$$S_j = \frac{1}{\left(\sum_{k=1}^{N_j} w_{k,j} \right) - 1} \sum_{k=1}^{N_j} w_{k,j} (X_k - M_j)(X_k - M_j)' \quad (6)$$

Um refinamento adicional na estimação dos graus de pertinência $w_{i,j}$ das amostras semi-rotuladas pela introdução da informação de contexto espacial extraída da imagem temática pode ser também investigado. Esse refinamento, o qual explora a correlação espacial existente entre as classes em uma cena natural, é conhecido como análise contextual. Uma descrição mais detalhada de métodos para utilização de informações de contexto em imagens digitais pode ser encontrada em Therrien (1989), entre outros. Neste estudo é empregada uma abordagem derivada de metodologia originalmente proposta por Di Zenzo *et al.* (1987) e Centeno (1991). O grau de pertinência de um pixel a uma classe é aqui estimado empregando um processo de relaxação probabilística contextual, o qual atualiza o grau de pertinência de cada pixel em função dos graus de pertinência que ocorrem nos pixels dentro de uma vizinhança.

Nesta proposta define-se como vizinhança, o conjunto dos oito pixels que circundam o pixel em consideração. O processo de relaxação probabilística faz uso do coeficiente de compatibilidade $r_{i,j}(h,k)$, que estima a compatibilidade de ocorrência simultânea da classe h em um pixel na posição i e da classe k em um vizinho na posição j . O coeficiente de compatibilidade aplicado nesse estudo é estimado pela equação (7) (Centeno, 1991):

$$r_{i,j} = \frac{1}{5} \ln \left(\frac{N(i,j,h,k) Ntotal}{\sum_{i=1}^m N(i,j,h,k) \sum_{j=1}^m N(i,j,h,k)} \right) \quad (7)$$

Onde $N(i,j,h,k)$ representa o número de ocorrências em que o pixel na posição i pertence à classe h e o seu vizinho na posição j pertence à classe k , m é o número de classes na imagem e $Ntotal$ representa o número total de pares de pixels na imagem. Os coeficientes de compatibilidade (7) são estimados a partir da imagem temática produzida a cada iteração na relaxação probabilística. Os graus de pertinência podem então ser atualizados a cada iteração conforme a equação (8) (Di Zenzo *et al.*, 1987):

$$w_{i,h}^{v+1} = \frac{1}{L} w_{i,h}^v \left(1 + \frac{1}{8} \sum_{j=1}^8 \sum_{k=1}^m r(i,j,h,k) w_{i,h}^v \right) \quad (8)$$

Onde $w_{i,h}^v$ representa o grau de pertinência do pixel na posição i , em relação à classe h , na iteração v , $w_{i,h}^{v+1}$ representa o grau de pertinência atualizado na iteração seguinte ($v+1$) e L representa um fator de normalização, dado pela equação (9):

$$L = \sum_{h=1}^m w_{i,h}^v \left(1 + \frac{1}{8} \sum_{j=1}^8 \sum_{k=1}^m r(i,j,h,k) w_{i,h}^v \right) \quad (9)$$

Na primeira iteração ($v=1$) adotam-se os valores estimados para os graus de pertinência obtidos na fase anterior, onde foram empregadas informações espectrais somente. Estes

valores são ajustados utilizando-se informações de contexto espacial (8) e novamente utilizados para uma ponderação mais acurada das amostras semi-rotuladas para uma nova estimação dos parâmetros do classificador nas equações (5) e (6).

3. Experimentos

Para testar a metodologia proposta desenvolveu-se um software em MATLAB, implementando os conceitos propostos na seção de Metodologia. Para fins de teste foram utilizados dados em alta dimensionalidade (imagem hiperespectral) obtida pelo sistema sensor AVIRIS sobre a área teste denominada de *Indian Pines*, localizada no noroeste do Estado de Indiana, USA (Figura 1). Esta cena compreende culturas de soja e milho, empregando técnicas de cultivo distintas (cultivo tradicional, cultivo direto, cultivo mínimo).

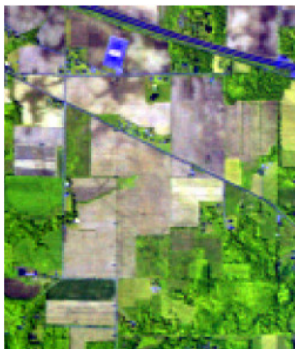


Figura 1a - Área Teste *Indian Pines*, Indiana, USA

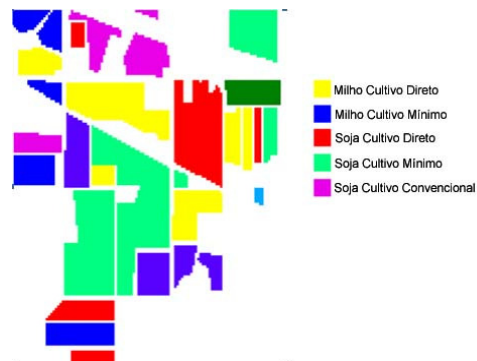


Figura 1b - Verdade Terrestre

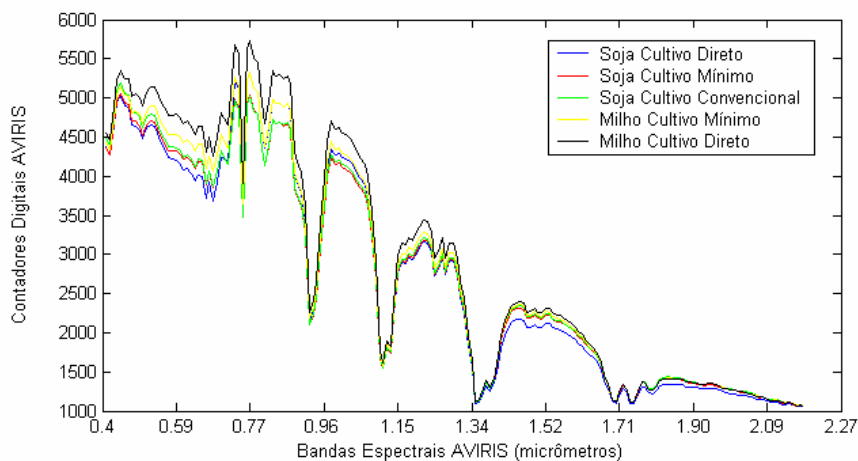


Figura 2 - Vetores de Médias das Classes Espectrais

Os dados AVIRIS foram coletados no início do cultivo, com uma substancial parcela do terreno mostrando solo exposto e resíduos de cultivos anteriores. Estas condições resultam em classes espectralmente muito similares entre si, isto é, com

vetores de médias muito próximos entre si (Figura 2). Nestas condições, a separabilidade entre as classes passa a depender basicamente das diferenças entre as matrizes de covariâncias.

Nestas circunstâncias, dados em alta dimensionalidade são necessários para obter-se uma imagem temática com um grau de acurácia satisfatório. Imagens temáticas produzidas a partir de dados multiespectrais tradicionais, como Landsat-TM por exemplo, resultam em valores muito baixos para a acurácia. O sistema sensor AVIRIS coleta dados em 224 bandas espectrais, no intervalo (0.4 – 2.5 μ m) do espectro eletromagnético (Landgrebe, 2003). Deste conjunto foram removidas bandas ruidosas (vapor de água na atmosfera), restando 190 bandas. Para fins deste estudo foram selecionadas cinco classes de cobertura do solo, todas espectralmente muito semelhantes entre si: milho cultivado direto, milho em cultivo mínimo, soja cultivado direto, soja em cultivo mínimo, soja em cultivo tradicional. Como ocorre frequentemente, o número de amostras disponíveis para fins de estimação dos parâmetros do classificador (amostras rotuladas) é limitado. Este fato impede uma estimação confiável dos parâmetros do classificador, especialmente na matriz de covariância, quando se utiliza somente as amostras rotuladas disponíveis. Desta forma, estes dados oferecem as condições necessárias à avaliação da metodologia proposta.

Os resultados dos experimentos realizados estão ilustrados na Figura 3. Das amostras de verdade terrestre disponíveis (Figura 1b) foram selecionados dois subconjuntos para cada uma das classes, um para fins de treinamento e outro para fins de teste (estimação da acurácia da imagem temática produzida). Para fins de geração da imagem temática foi utilizada a função de decisão Máxima Verossimilhança Gaussiana (MVG) em sua forma logarítmica (1), em três configurações distintas: (i) treinamento da função de decisão empregando apenas amostras rotuladas (Método 1), (ii) adição de amostras semi-rotuladas ponderadas por informações espectrais somente (Método 2), e (iii) idêntico ao anterior, com a ponderação por informações espectrais seguida de ponderação por contexto espacial (Método 3). Em cada caso, a acurácia da imagem foi estimada por meio das amostras de teste, para diferentes valores da dimensionalidade dos dados (bandas espectrais), conforme ilustrado na Figura 3.

Neste experimento foram empregadas 100 amostras rotuladas, em seguida acrescidas com 50, 100 e 200 amostras semi-rotuladas. Na legenda da Figura 3, estes experimentos são denotados por “Método 2 (50)”, “Método 2 (100)”, “Método 2 (200)”, “Método 3 (50)”, “Método 3 (100)” e “Método 3 (200)”. Nestes casos, por exemplo, o “Método 3 (200)” é aquele que faz uso da informação de contexto espacial acrescido ao método de semi-rotulação, com 200 amostras semi-rotuladas. Os valores estimados para a acurácia média das cinco classes estão ilustrados na Figura 3 para as dimensões iguais a 20, 40, 60 e 80 bandas espectrais. Os resultados, conforme ilustrados naquela figura tendem a confirmar a adequação da metodologia proposta neste estudo (Método 3). O experimento implementando o Método 1, utiliza somente as 100 amostras rotuladas disponíveis, um número pequeno quando comparado com a dimensionalidade dos dados. Esta limitação no número de amostras resulta no efeito de Hughes que pode ser claramente observado na Figura 4 quando a dimensão dos dados cresce. A adição de amostras semi-rotuladas ponderadas por informações espectrais somente (Método 2) resulta em um visível incremento no valor estimado para a acurácia na dimensionalidade

limite, ou seja, quando 100 amostras rotuladas são utilizadas na dimensão 80. Os resultados obtidos pelo Método 2, empregando 50, 100 e 200 amostras semi-rotuladas na dimensionalidade mais alta comprova que o uso desse método é eficaz quando o número inicial de amostras rotuladas é muito próximo à dimensão dos dados. Pode-se observar uma sensível redução nos efeitos do fenômeno de Hughes com o número crescente de amostras semi-rotuladas inseridas no processo. Finalmente, foram incluídas no experimento amostras semi-rotuladas ponderadas conforme o Método 3. Os resultados ilustram claramente a contribuição da informação de contexto espacial na ponderação das amostras semi-rotuladas. Em todos os três experimentos (adição de 50, 100 e 200 amostras semi-rotuladas) os valores estimados para a acurácia foram claramente superiores aos obtidos pelos demais processos. O resultado superior do Método 3 destaca-se inclusive nas dimensões menores, onde o uso do Método 2 não apresenta grande contribuição à acurácia final, em relação ao uso exclusivo de amostras rotuladas (Método 1).

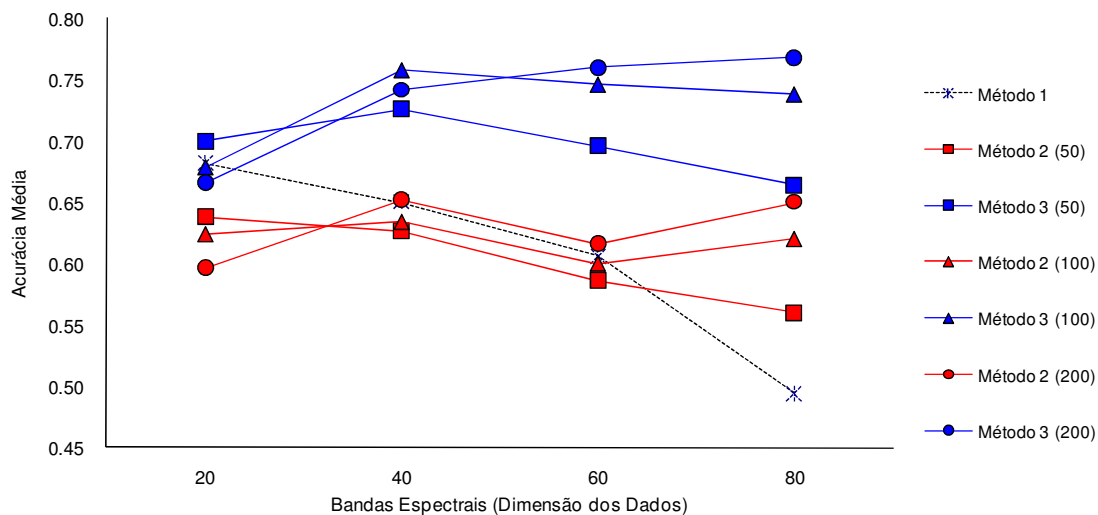


Figura 3 – Acurácia média de classificação para as cinco classes, três métodos e diferentes quantidades de amostras semi-rotuladas

4. Conclusão

Neste estudo foi investigada a contribuição das informações de contexto espacial na ponderação de amostras semi-rotuladas. Amostras semi-rotuladas se constituem em um recurso eficiente e prático para incrementar o número de amostras utilizadas para fins de treinamento em classificadores paramétricos. Na literatura são encontradas propostas de metodologias para fins de ponderação destas amostras. Neste estudo é proposta e testada uma nova abordagem fazendo uso de dados de contexto espacial em uma imagem temática. A metodologia investigada parte da abordagem proposta por Jackson e Landgrebe (2001). Naquela metodologia a ponderação das amostras semi-rotuladas é realizada em um processo iterativo, empregando informações espectrais extraídas no próprio processo de classificação. Na metodologia proposta neste estudo, os graus de pertinência associados a cada uma das amostras semi-rotuladas sofrem um processo

adicional de refinamento, empregando informações de contexto espacial. A metodologia proposta foi testada empregando dados hiperespectrais coletados pelo sistema sensor AVIRIS cobrindo uma área teste com classes espectralmente muito similares entre si. Os experimentos, empregando diferentes valores para a dimensionalidade dos dados e para número de amostras semi-rotuladas mostraram claramente a contribuição da informação de contexto na ponderação das amostras semi-rotuladas. Experimentos adicionais são, entretanto, necessários para um entendimento mais completo do comportamento da metodologia aqui proposta.

5. Referencias Bibliográficas

Fukunaga, K. **Introduction to Statistical Pattern Recognition**. 2nd edition, Academic Press, 1990.

Landgrebe, D. **Signal Theory Methods in Multispectral Remote Sensing**. Wiley-Interscience, 2003.

Shashahani, B.M.; Landgrebe, D. The effect of Unlabeled Samples in Reducing the Small Sample Size Problem and Mitigating the Hughes Phenomenon. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 32, No. 5, pp. 1087-1095, 1994.

Jackson, Q.; Landgrebe, D. An Adaptive Classifier Design for High-Dimensional Data Analysis With a Limited Training Data Set. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 39, No. 12, pp. 2664-2679, 2001.

Liczbinski, C.; Haertel, V., A New Approach to Estimate A Priori Probabilities in Remote Sensing Digital Image Classification, *Canadian Journal of Remote Sensing*, Vol. 34, No. 2, pp. 135-142, 2008.

Therrien, C. W., **Decision Estimation and Classification, An Introduction to Pattern Recognition and Related Topics**, John Wiley & Sons, New York, 1989.

Di Zeno, S., Bernstein, R., DeGloria, S., Kolsky, H. G., Gaussian Maximum Likelihood and Contextual Classification Algorithms for Multicrop Classification, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. GE-25, No. 6, pp. 805-814, 1987.

Centeno, J. A. S., **Exploração de Atributos de Contexto em Imagens Digitais no Estudo da Cobertura Vegetal em Bacias Hidrográficas**, Dissertação de Mestrado, Programa de Pós-Graduação em Engenharia Civil, Área de Concentração Recursos Hídricos e Saneamento, UFRGS, 1991.