

Estimação de controladores de influência em classificadores multidimensionais com abordagem Bayesiana

Orlando Alves Máximo^{1,2}
David Fernandes²

¹Instituto de Estudos Avançados - IEAv/CTA
Rodovia dos Tamoios Km 5,5 - 12228-001 - São José dos Campos - SP, Brasil
oamaximo@ieav.cta.br

²Instituto Tecnológico de Aeronáutica - ITA/CTA
Praça Marechal Eduardo Gomes, 50 - 12 228-900 - São José dos Campos - SP, Brasil
{maximo, david}@ita.br

Abstract. This paper evaluates the accuracy of Bayesian classifiers using reliability factors estimated by the entropy of image inter-class Bhattacharyya distance. The results show that the use of reliability factors improves the accuracy of the classifier in comparison with a standard Bayesian classifier (the Cascade Classifier in this case). When compared to exhaustive evaluation, the entropy estimated reliability factors lies within 3% from the best results.

Palavras-Chaves: remote sensing, classification, SAR, sensoriamento remoto, processamento de imagens, classificadores, função de adesão, entropia.

1. Introdução

O interesse da comunidade científica internacional no desenvolvimento de técnicas de análise de imagens de sensoriamento remoto tem crescido bastante nas últimas décadas. Diversas razões são apontadas para justificar tal cenário. Uma delas é o caráter interdisciplinar inerente às imagens de sensoriamento remoto, o que as torna atraentes para utilização em várias aplicações, tais como: cartografia, hidrologia, planejamento urbano, agricultura, prevenção e monitoramento de catástrofes, preservação e recuperação do meio-ambiente, vigilância territorial e operações de reconhecimento em missões militares.

Outro fator que contribuiu para a manutenção do interesse nas imagens de sensoriamento remoto foi a contínua melhoria nas características dos sensores imageadores e das plataformas nas quais estes sensores estão instalados. Atualmente, estão disponíveis imagens de alta resolução geradas por sensores que operam tanto na faixa óptica quanto na faixa de microondas, com capacidade de imageamento multicanal (multi ou hiperespectral no caso dos sensores ópticos e também multifrequência e multipolarimétricos no caso dos sensores de microondas).

Para uma exploração mais efetiva do potencial destas imagens, são utilizadas cada vez mais as técnicas automáticas de extração de informação apoiadas em processos computacionais. Uma das técnicas mais utilizadas é a classificação, que consiste em atribuir a cada pixel da imagem uma classe, dentro de um determinado conjunto de classes pré-estabelecido (Campbell, 1996).

Neste contexto, a abordagem Bayesiana fornece uma sólida base para a construção de classificadores (Devjiver e Kitter, 1982) e para a minimização de erros (Fukunaga, 1990), mesmo quando o problema de classificação envolve a definição de densidades de probabilidades multidimensionais. Dentro desta abordagem, um dos classificadores mais utilizados é o Classificador de Máxima Verossimilhança (MaxVer). Outro exemplo que pode ser citado é o Classificador em Cascata (Swain, 1978)

Nestes classificadores, cada dimensão é gerada por uma imagem representando uma banda (no caso de sensores multi ou hiperespectrais), uma data (no caso de imagens multitemporais), uma polarização (no caso de imagens multipolarizadas) etc.

Nestas situações, entretanto, os classificadores construídos a partir da abordagem Bayesiana assumem o pressuposto de que todas as dimensões (ou, simplesmente, as imagens) exercem a mesma influência no resultado final da classificação, independentemente da qualidade das mesmas.

Uma maneira de ponderar a influência das diferentes imagens consiste em introduzir “controladores de influência” para cada imagem utilizada na classificação. Um exemplo desta abordagem é a Função de Adesão Modificada - MGMF (Benediktsson, Swain e Ersoy, 1990).

O objetivo dos controladores de influência é expressar quantitativamente o grau de participação que cada imagem pode ter no processo de classificação. Desta forma, a influência das imagens mais relevantes é realçada e a influência das imagens menos relevantes é minimizada, com o intuito de melhorar o resultado final da classificação.

Neste contexto, o problema do critério de escolha dos melhores controladores e a metodologia para sua estimação constitui-se num campo de pesquisas promissor.

Assim, este trabalho tem como objetivo apresentar os resultados da utilização de uma metodologia de estimação de controladores de influência baseada na medida da entropia do conjunto de distâncias de Bhattacharyya entre as classes presentes na imagem.

2. Controladores de Influência e a Função de Adesão

A abordagem Bayesiana presente no Classificador MaxVer e no Classificador em Cascata vem sendo usada há décadas para o desenvolvimento e a implementação de técnicas para classificação assistida por computador.

Como comentado anteriormente, nos casos multidimensionais, percebe-se que a abordagem Bayesiana apresenta a característica de conferir a todas as dimensões a mesma capacidade de influência no resultado final da classificação. Este fato, aliado ao caráter multiplicativo presente nas funções discriminantes dos classificadores, introduz uma possibilidade de erro significativa no processo de classificação. Se uma das imagens (dimensões) apresentar alguma anomalia, ela irá contaminar todo o processo, pois não há um mecanismo para o isolamento ou a minimização da influência desta imagem.

Cada membro da multiplicação representa uma probabilidade com valores no intervalo $[0,1]$. Se alguma delas assumir valores próximos a zero, o resultado final da função discriminante será próximo de zero, mascarando a contribuição das outras imagens, mesmo que a observação em questão tenha pouca ou nenhuma qualidade (Benediktsson e Swain, 1989).

Torna-se desejável, então, a introdução de fatores de ponderação para as observações que controlem (ou influenciem) a sua contribuição para o processo de classificação. Estes fatores de ponderação devem atuar de modo a minimizar a influência das observações com menor qualidade, preservando a influência daquelas que apresentam melhores características, com o objetivo de aumentar a precisão do processo de classificação.

Dentro deste enfoque de utilização de pesos controladores de influência das observações num processo de classificação, pode ser citada a ‘Função de Adesão’ (Lee, Richards e Swain, 1987) que define os controladores como graus de confiança ou confiabilidade das observações (imagens).

Considerando um conjunto de n observações independentes, cada uma fornece uma medida x_k , com $k = 1, \dots, n$ para um determinado pixel de interesse.

Seja $W = \{w_1, \dots, w_m\}$ o conjunto pré-definido de classes dentro do qual os pixels deverão ser classificados.

Partindo destas premissas, Lee, Richards e Swain (1987) estabeleceram a Função de Adesão para cada classe e observação como sendo $f(w_j|x_k)$. Esta função mede o nível de vinculação ou associação entre a classe w_j e a medida específica de uma observação, x_k .

A partir das funções de adesão de cada observação, é possível estabelecer uma Função de Adesão Global (*Global Membership Function* - GMF) que concatene as informações provenientes de todas as observações referentes a uma classe específica. Para esta concatenação, ainda segundo Lee, Richards e Swain (1987), é prudente associar a cada observação uma grandeza que expresse a confiança na informação proveniente dela.

A GMF pode ser utilizada como função discriminante num processo de classificação. Desta forma, o vetor de atributos X é classificado de acordo com a regra usual de seleção do maior valor, ou seja:

$$X \in w_j \Leftrightarrow F_j(X) = \max_i F_i(X) . \quad (1)$$

Para a implementação deste classificador, é necessário definir explicitamente a Função de Adesão $F_j(\cdot)$ que combina $f(w_j|x_k)$ e o conjunto de graus de confiança, um para cada observação x_k , $k = 1, \dots, n$.

Utilizando-se a abordagem Bayesiana, uma maneira possível para o controle de influência das imagens é a adição de expoentes às probabilidades *a posteriori* de cada observação. Deste modo, a Função de Adesão do pixel com vetor de atributos X pode ser escrita como (Lee, Richards e Swain, 1987):

$$F_j(X) = p(w_j)^{1-n} \prod_{k=1}^n p(w_j|x_k)^{\alpha_k} \quad (2)$$

onde α_k controla a influência da observação k e está no intervalo $[0,1]$.

Outro modo de se atribuir os graus de confiança (ou grau de qualidade) foi proposto por Benediktsson e Swain (1989), que analisaram a influência de cada observação dentro da GMF. Para isso, calcularam a razão entre a Função de Adesão para $X = \{x_1, \dots, x_n\}$ e para $X = \{x_1, \dots, x_{n+1}\}$, dadas, respectivamente, por:

$$F_j(x_1, x_2, \dots, x_n) = p(w_j)^{1-n} \prod_{k=1}^n p(w_j|x_k). \quad \text{e} \quad (3)$$

$$F_j(x_1, x_2, \dots, x_{n+1}) = p(w_j)^{-n} \prod_{k=1}^{n+1} p(w_j|x_k). \quad (4)$$

Calculando-se a razão entre a **Equação (3)** e a **Equação (4)** obtém-se a influência da observação adicional, que é dada por:

$$\frac{F_j(x_1, x_2, \dots, x_{n+1})}{F_j(x_1, x_2, \dots, x_n)} = \frac{p(w_j)^{-n} \prod_{k=1}^{n+1} p(w_j|x_k)}{p(w_j)^{1-n} \prod_{k=1}^n p(w_j|x_k)} = \frac{p(w_j|x_{n+1})}{p(w_j)}. \quad (5)$$

Deste modo:

$$F_j(x_1, x_2, \dots, x_{n+1}) = F_j(x_1, x_2, \dots, x_n) \frac{p(w_j|x_{n+1})}{p(w_j)},$$

e $p(w_j|x_{n+1})/p(w_j)$ representa a contribuição da observação x_{n+1} na Função de Adesão.

A expressão definida pela **Equação (3)** pode ser reescrita de modo a salientar a influência de cada observação na Função de Adesão da seguinte forma:

$$F_j(x_1, x_2, \dots, x_n) = p(w_j) \prod_{k=1}^n \frac{p(w_j|x_k)}{p(w_j)}.$$

Assim, para controlar a influência de cada observação dentro do processo de classificação, os controladores de influência α_i são definidos como expoentes das contribuições individuais de cada observação, criando a Função de Adesão Modificada (*Modified Global Membership Function* - MGMF), $F'_j(X)$, que é dada por:

$$F'_j(x_1, x_2, \dots, x_n) = F'_j(X) = p(w_j) \prod_{k=1}^n \left[\frac{p(w_j|x_k)}{p(w_j)} \right]^{\alpha_k}, \quad (6)$$

onde os α_k 's são selecionados no intervalo $[0,1]$.

Desta forma, se uma fonte é totalmente não-confiável, seu controlador de influência é zero e ela não terá nenhuma influência no resultado final da classificação, independentemente do valor de $p(w_j|x_k)$. No extremo oposto, caso a fonte seja totalmente confiável, seu controlador de influência será 1 e seu impacto no resultado final da classificação será totalmente preservado.

A regra de decisão para a alocação de uma classe ao pixel com vetor de atributos X fica:

$$X \in w_j \Leftrightarrow F'_j(X) = \max_i F'_i(X). \quad (7)$$

É conveniente ressaltar que a técnica de introduzir controladores de influência na forma de expoentes não afeta o resultado para um processo de classificação com uma única observação devido à monotonicidade de $F'_j(x_1)$.

A expressão definida na **Equação (6)** também pode ser reescrita na forma logarítmica como:

$$\log F'_j(X) = \log p(w_j) + \sum_{k=1}^n \alpha_k \log \frac{p(w_j|x_k)}{p(w_j)} \quad (8)$$

Máximo e Fernandes (2003) propuseram ainda controladores de influência alocados a cada uma das classes presentes em cada uma das observações, generalizando ainda mais o conceito de controladores de influência.

3. Estimadores de Controladores de Influência

O uso dos controladores de influência a serem atribuídos às imagens impõe a solução do problema da quantificação da qualidade de cada imagem e a posterior definição dos valores dos controladores (α_k) baseados nesta quantificação.

Uma escolha natural é associar pesos comparativamente maiores para as “melhores” imagens e pesos menores para as “piores”. Infelizmente, “**melhor**” e “**piores**” são definições subjetivas e afirmar que uma imagem é duas vezes melhor (ou pior) que outra é apenas uma figura de linguagem. Sem uma especificação precisa das características das imagens que devem ser levadas em consideração, o problema fica mal definido e torna-se evidente uma falta de consistência nos resultados (French, 1981).

Uma estratégia para escolha dos controladores pode ser baseada na separabilidade das classes presentes na imagem, uma vez que um aumento da distância entre as distribuições das classes corresponde a uma diminuição do Erro de Bayes para o processo de classificação (Fukunaga, 1990).

Neste contexto, ainda segundo Fukunaga (1990), a Distância de Bhattacharyya apresenta uma relação direta com a diminuição do Erro de Bayes.

A distância média entre as classes modeladas nas imagens poderia, então, ser utilizada como parâmetro para a avaliação das mesmas e para a alocação do seu peso. As imagens que apresentarem uma distância média maior são consideradas mais ‘confiáveis’ por apresentarem uma probabilidade de erro menor e, dessa forma, recebem pesos maiores.

Ocorre, porém, que a utilização direta do valor médio das distâncias entre as classes não leva em consideração as relações entre as mesmas, mascarando, por exemplo, a existência de subconjuntos de classes estatisticamente próximas.

Para uma análise mais apurada do relacionamento entre as distâncias inter-classes, é necessário levar em consideração a organização das mesmas em relação ao padrão ideal, que seria a equidistância entre todas as classes. Esta situação ideal reflete a melhor utilização do espaço de atributos pelas classes. Por exemplo: O vetor de distâncias inter-classes $D_1 = \{d_{12}; d_{13}; d_{23}\}$ representado por $D_1 = \{0, 8; 0, 8; 0, 2\}$, evidencia a existência de duas classes estatisticamente próximas, o que dificulta a discriminação entre elas. Apesar de possuir a mesma média que o vetor D_1 , um vetor $D_2 = \{0, 6; 0, 6; 0, 6\}$ evidencia uma melhor separabilidade entre as classes na imagem.

As imagens podem, então ser classificadas de acordo com o afastamento desta situação ideal. Quanto mais próximo do ideal, maior o grau de confiança (controlador de influência) atribuído àquela imagem.

A medida deste afastamento pode ser obtida através da entropia (H_k) do vetor de distâncias entre m classes para a observação k , definida por:

$$H_k = - \sum_{i=1}^{m-1} \sum_{j=i+1}^m \delta_{ij} \log_{\frac{m^2-m}{2}} \delta_{ij} \quad (9)$$

onde $0 \leq H_k \leq 1$.

$$\delta_{ij} = \frac{d_{ij}}{\sum_{i=1}^{m-1} \sum_{j=i+1}^m d_{ij}} \leq 1 \quad (10)$$

e d_{ij} é a distância de Bhattacharyya entre as classes i e j da observação.

Se $\delta_{ij} = 2/(m^2 - m) \forall i, j$, ou seja, as distâncias entre as classes são sempre as mesmas, $H_k = 1$.

Se $\delta_{ij} = 1$ para um dado par de classes i, j , então $H_k = 0$.

Este valor de entropia será associado diretamente ao grau de confiança da observação k , ou seja $\alpha_k = H_k$.

4. Descrição do Experimento

Para a avaliação da utilização da entropia do vetor de distâncias inter-classes como estimador dos controladores de influência das imagens, foi construído um classificador baseado na Função de Adesão Modificada (**Equação 6**), com a regra de decisão definida na **Equação 7**.

Foi selecionado um conjunto de dados SAR multifreqüência, composto por duas imagens da região de Freiburg, na Alemanha (uma na banda X e outra na banda L) de 512 x 512 pixels. Estas imagens foram obtidas pelo sensor aerotransportado SAR-580. As **Figuras 1 e 2** apresentam as duas imagens originais utilizadas no trabalho.

Foi definido um conjunto de 4 classes nas imagens, que foram modeladas utilizando-se uma densidade de probabilidade gaussiana. Para uma melhor adequação das características estatísticas das imagens, foi realizado um pré-processamento com filtros da média com janelas 3x3, 5x5, 7x7, 9x9 e 11x11, totalizando 6 conjuntos de dados, um deles constituído pelas imagens originais (janela 1x1).

Após o pré-processamento, foram selecionadas amostras nas classes de referência e realizada a modelagem das mesmas. Os parâmetros obtidos (média e variância) foram utilizados para o cálculo dos vetores de distâncias inter-classes para cada uma das imagens.

Como comentado anteriormente, para a definição da separabilidade das classes, foi utilizada a Distância de Batthacharyya definida em Fukunaga (1990), e a partir dessas distâncias foi

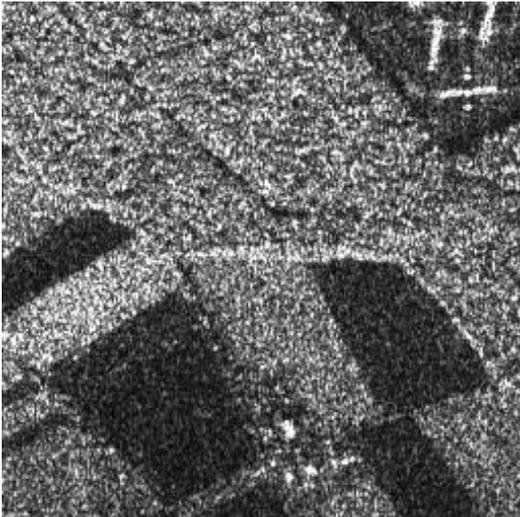


Figura 1. Imagem na Banda X



Figura 2. Imagem na Banda L

calculada a entropia para cada conjunto de distâncias de cada imagem, estabelecendo-se assim, o grau de influência de cada uma.

Para que pudesse ser feita uma comparação, os conjuntos de imagens originais e pré-processadas com os filtros da média também foram classificados com o classificador em cascata definido por Swain (1978).

A precisão dos classificadores foi obtida através do Coeficiente Kappa.

Para a extração da matriz de erro, foi utilizada a técnica de amostragem AESNA (Amostragem Estratificada Sistemática Não-Alinhada). Esta técnica apresenta a característica de fornecer uma densidade de pontos de amostragem uniforme em todas as regiões da imagem (Fidalgo, 1995). Foram utilizados 441 pontos de amostragem.

Como comentado anteriormente, o objetivo da utilização dos controladores de influência é aumentar o desempenho do classificador. Para que se pudesse avaliar a qualidade dos controladores obtidos a partir da entropia do vetor de distâncias inter-classes, eles foram comparados com os controladores que maximizam o desempenho do classificador.

Para a definição destes controladores de máximo desempenho, foi realizada uma avaliação exaustiva, que consistiu na classificação das imagens variando-se os valores dos controladores de cada uma das imagens no intervalo $[0,1]$, com incrementos de 0,01, totalizando 10.000 classificações para cada conjunto.

5. Resultados e Discussão

De maneira geral, o classificador baseado na Função de Adesão Modificada e utilizando os controladores de influência baseados na entropia do vetor de distâncias inter-classes (MGMF-entropia) apresentou desempenho superior ao Classificador em Cascata em todos os conjuntos de imagens, conforme pode ser observado na **Figura 3**.

Nota-se que há um incremento no desempenho dos dois classificadores em função do aumento da janela de filtragem, com uma tendência de saturação a partir da janela 11x11.

A diferença de desempenho entre os classificadores também tende a reduzir com o aumento da janela de filtragem.

Na **Figura 4**, mostra-se o desempenho da classificação para o conjunto de imagens pré-processados com filtro da média com janela 5x5, variando-se os valores dos controladores de influência e utilizando-se o parâmetro 'Distância de Precisão' como métrica de desempenho. A Distância de Precisão foi definida em Máximo e Fernandes (2002) e utiliza cinco métricas

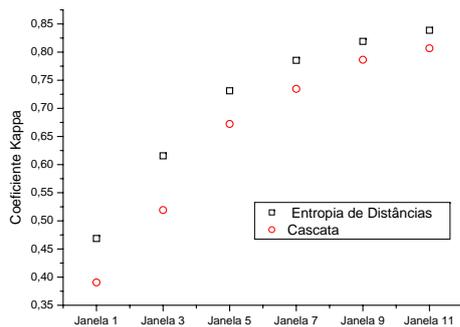


Figura 3. Comparação entre o Classificador MGMF-entropia e o Classificador em Cascata para várias janelas de filtragem.

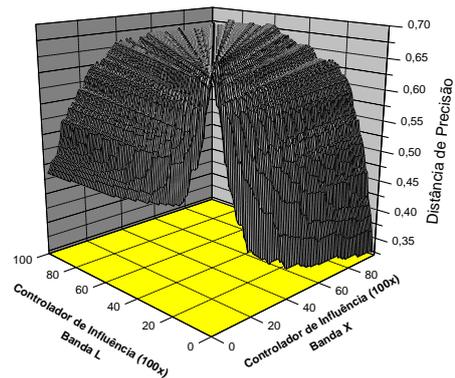


Figura 4. Precisão do Classificador MGMF em função da combinação de controladores de influência, utilizando-se um filtro da média com janela 5x5.

de avaliação de desempenho (Precisão Global, Precisão Média por Classes, Coeficiente Kappa, Distância de Kullbak-Leibler Proporcional e Distância de Kullback-Leibler Uniforme). Pela análise da figura constata-se que não há uma combinação única de controladores de influência que maximiza o desempenho.

Comparando-se numericamente o resultado da classificação com os controladores de influência estimados pela entropia, que resultou nos valores $\alpha_X = 0,752$ e $\alpha_L = 0,832$, e a avaliação exaustiva representada na **Figura 4**, constata-se que os primeiros estão em uma faixa de tolerância de 3% em relação aos valores máximos de desempenho obtidos.

As **Figuras 5 e 6** apresentam o resultado da classificação com os controladores de influência calculados pela entropia para o conjunto de imagens originais e pré-processadas com o filtro da média com janela 5x5.

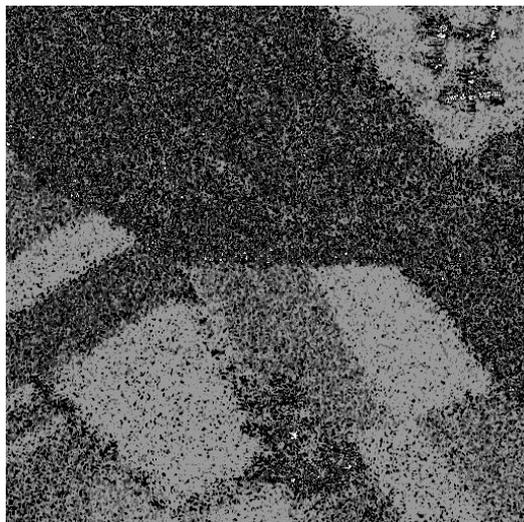


Figura 5. Resultado da classificação com o classificador MGMF-entropia para o conjunto de imagens originais, com $\alpha_X = 0,752$ e $\alpha_L = 0,832$.

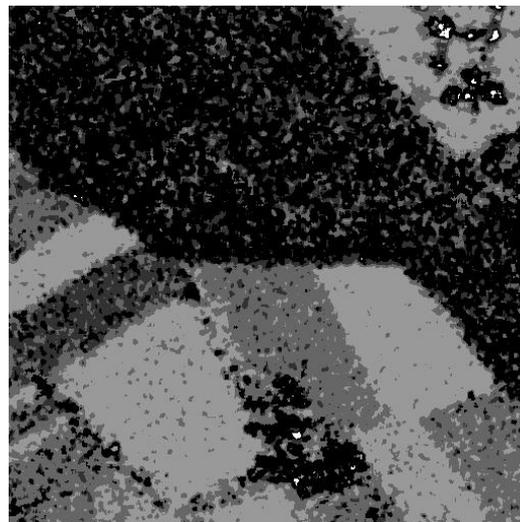


Figura 6. Resultado da classificação com o classificador MGMF-entropia, imagens pré-processadas com filtro da média, janela 5x5, com $\alpha_X = 0,741$ e $\alpha_L = 0,831$.

6. Conclusões

Apresentou-se neste trabalho uma metodologia para estimação dos controladores de influência de imagens utilizadas em um processo de classificação Bayesiano. A estimativa destes controladores

para uma imagem consiste no cálculo da entropia dos valores de distância de Bhattacharyya, normalizados, entre as classes de alvos presentes na imagem. Os controladores de influência estimados podem ser utilizados no GMF ou na MGMF. Outras medidas de distância inter-classes também podem ser utilizadas.

A estimativa dos controladores de influência para duas imagens do sensor SAR-580 apresentou um desempenho na faixa de tolerância de 3% do melhor desempenho segundo o critério da Distância de Precisão (Máximo e Fernandes, 2002), evidenciando que o processo de estimação ainda pode ser melhorado.

Referências

Benediktsson, J. A.; Swain, P. H. A method of statistical multisource classification with a mechanism to weight the influence of the data sources. In: International Geoscience and Remote Sensing Symposium - IGARSS89. **Proceedings...** Vancouver: IEEE, 1989. p. 517-520. CD-ROM.

Benediktsson, J. A.; Swain, P. H.; Ersoy, O. K. Neural network approaches versus statistical methods in classification of multisource remote sensing data. **IEEE Transactions on Geoscience and remote sensing**, v. 28, n. 4, p. 540-552, July 1990.

Campbell, J. B. **Introduction to remote sensing**. 2. ed. New York: Guilford Press, 1996. 622 p.

Devjiver, P. A.; Kitter, J. **Pattern recognition: A statistical approach**. London: Prentice Hal, 1982. 448 p.

Fidalgo, E. C. C. **Exatidão no processo de mapeamento temático da vegetação de uma área de mata atlântica no estado de São Paulo, a partir de imagens TM-Landsat**. 1995. (INPE-5944-TDI/570). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1995.

French, S. Consensus of opinion. **European Journal of Operational Research**, v. 7, n. 4, p. 332-340, Aug. 1981.

Fukunaga, K. **Introduction to statistical pattern recognition**. 2. ed. San Diego: Academic Press, 1990. 591 p.

Lee, T.; Richards, J.; Swain, P. H. Probabilistic and evidential approaches for multisource data analysis. **IEEE Transactions on Geoscience and remote sensing**, GE-25, n. 3, p. 283-293, May 1987.

Máximo, O. A.; Fernandes, D. The use of data set reliability factors in a multidimensional Bayesian classifier. **International Geoscience and Remote Sensing Symposium (IGARSS - 2002)**, Toronto, Canada, June 24-28.

Máximo, O.A.; Fernandes, D. Uso de graus de confiança das classes em classificadores Bayesianos. In: Simpósio Brasileiro de Sensoriamento Remoto, 11., 2003, Belo Horizonte. **Anais...** São José dos campos: INPE, 2003. p. 2083-2090. Disponível na biblioteca digital URLib: <http://tid.inpe.br/sbsr/2002/11.17.11.53>.

Swain, P. H. Bayesian classification in a time-varying environment. **IEEE Trans. Sys. Man. And Cyber**, SMC-8, n. 12, p. 879-883, Dec. 1978.