

APLICAÇÃO DE UM MÉTODO FUNDAMENTADO EM GRAFO E DISTÂNCIAS ESTOCÁSTICAS NA CLASSIFICAÇÃO BASEADA EM REGIÕES

Rogério Galante Negri¹
Luciano Vieira Dutra²
Tatiana Sussel Gonçalves Mendes¹

¹ Universidade Estadual Paulista – UNESP
Instituto de Ciência e Tecnologia – Departamento de Engenharia Ambiental
Rodovia Presidente Dutra, km 137,8 – 12247-004 – São José dos Campos – SP, Brasil
{rogerio.negri, tatiana.mendes}@ict.unesp.br

² Instituto Nacional de Pesquisas Espaciais – INPE
Divisão de Processamento de Imagens
Caixa Postal 515 – 12245-970 – São José dos Campos – SP, Brasil
dutra@dpi.inpe.br

Abstract. Traditionally, classification of remote sensing images has been performed using only the spectral pixels' information through the so-called Pixel Based Methods. However, there are cases which the pixel based approach is not able to provide adequate results. An alternative is the use of Region Based Classification. The region based classification can be performed in different ways, for example, by methods based on minimum distance or Support Vector Machine (SVM). This work investigated the application of Graph Classification for region based classification. A case study about the land use and land cover classification on an Amazon region adopting the SPOT image was conducted. The comparison made between Graph Classification, SVM and the Minimum Distance show that Graph Classification achieves better results.

Keywords: classificação baseada em regiões, distância estocástica, função *kernel*

1. Introdução

Tradicionalmente, a classificação de imagens de sensoriamento remoto tem sido realizada utilizando apenas a informação espectral pixels, comumente denominada Classificação Baseada em Pixels. No entanto, devido a forte heterogeneidade das imagens de alta resolução, o uso somente da informação espectral pode ser insatisfatória (GIGANDET et al., 2005). Uma alternativa neste caso é o uso da Classificação Baseada em Regiões, onde primeiramente os pixels são agregados em objetos homogêneos através de técnicas de segmentação de imagens, e em seguida, estes objetos são classificados individualmente (DESHENG; FAN, 2010).

A Classificação Baseada em Regiões pode ser realizada de diferentes maneiras. Um exemplo é a abordagem apresentada em Silva et al. (2011), onde o Classificador de Mínima Distância Estocástica é adotado. O emprego de Máquina de Vetores Suporte (SVM - *Support Vector Machine*) para Classificação Baseada em Regiões é verificada em Negri, Dutra e Sant'Anna (2012) através do emprego da função *kernel* Bhattacharyya, a qual também é fundamentada em conceitos de distâncias estocásticas. Seguindo a mesma motivação apresentada em Negri, Dutra e Sant'Anna (2012), verifica-se que o método de Classificação por Grafo, proposto em Camps-Valls, Tatyana e Zhou (2007), inicialmente destinada à Classificação Baseada em Pixels, pode ser empregada na Classificação Baseada em Regiões a partir da utilização da função *kernel* Bhattacharyya.

Dessa forma, este estudo apresenta a utilização da Classificação por Grafo no problema de Classificação Baseada em Regiões, comparando por sua vez os resultados obtidos com as

propostas apresentadas em Silva et al. (2011) e Negri, Dutra e Sant'Anna (2012). Para tal, um estudo de caso sobre a classificação do uso e cobertura do solo em uma região no entorno da Floresta Nacional do Tapajós com uso de imagem do satélite SPOT foi realizado.

2. Fundamentação Teórica

2.1. Classificação Baseada em Regiões e o Classificador de Mínima Distância Estocástica

Seja \mathcal{I} uma imagem definida sobre um suporte $\mathcal{S} \subset \mathbb{N}^2$ e \mathcal{X} o espaço de atributos, $\mathcal{I}(s) = \mathbf{x}$ denota que o pixel $s \in \mathcal{S}$ de \mathcal{I} possui vetor de atributos $\mathbf{x} \in \mathcal{X}$. O processo de Classificação Baseada em Regiões consiste em associar a classe ω_j , $j = 1, \dots, c$ à região $R_i \subset \mathcal{S}$, $i = 1, \dots, r$. R_i é um conjunto de pixels s_a , $a = 1, \dots, \#R_i$, onde os atributos de s_a são obtidos de $\mathcal{I}(s_a)$ e $\#$ é o operador de cardinalidade. Neste contexto, o suporte de \mathcal{I} é particionado em r regiões disjuntas através de um processo de segmentação. Assim, as regiões representam conjuntos de pixels espacialmente conectados, os quais atendem determinado critério de uniformidade. No processo de classificação, todos os pixels de uma mesma região são associados a uma única classe.

Para os métodos de classificação baseados em região e de aprendizado supervisionado, é necessário o emprego de um conjunto de regiões rotuladas $\mathcal{D}_l = \{(R_i, \omega_j) \in \mathcal{S} \times \Omega : i = 1, \dots, m; j = 1, \dots, c\}$, onde m é o número de regiões de treinamento. Por fim, a notação (R_i, ω_j) indica que R_i está associada a classe ω_j .

Em Silva et al. (2011) o Classificador de Mínima Distância Estocástica é empregado para Classificação Baseada em Regiões. O processo de classificação definido por este método consiste em estimar uma distribuição de probabilidade para cada uma das regiões não rotuladas e, em seguida, associar tal região à classe que possui distribuição de probabilidade mais semelhante. A distribuição de probabilidade das classes que abrangem o problema de classificação são estimadas a partir das informações presentes em \mathcal{D}_l , já a medida de proximidade entre distribuições de probabilidade é determinada via Distâncias Estocásticas, como o caso da distância de Bhattacharyya, que sobre consideração de distribuição Gaussiana Multivariada dos dados, é definida por (RICHARDS; XIUPING, 2005):

$$\mathbf{B}_G(R_c, R_d) = \frac{1}{8} (\mu_c - \mu_d)^T \left(\frac{\Sigma_c + \Sigma_d}{2} \right)^{-1} (\mu_c - \mu_d) + \frac{1}{2} \ln \left(\frac{|0.5 (\Sigma_c + \Sigma_d)|}{\sqrt{|\Sigma_c| + |\Sigma_d|}} \right), \quad (1)$$

onde μ_z e Σ_z são o vetor média e a matriz de covariância, estimadas a partir das informações da região R_z . Em (1), $(\cdot)^T$, $|\cdot|$ e $(\cdot)^{-1}$ representam as operações de transposição, determinante e inversão matricial, respectivamente.

Formalmente, dada a região não classificada R_i e sendo $\mathbf{B}_G(f_{R_i}, f_{\omega_j})$ a distância estocástica entre as distribuições de probabilidade obtidas dos pixels de R_i e da classe ω_j , o Classificador de Mínima Distância Estocástica realiza a associação (R_i, ω_j) baseado na seguinte regra:

$$(R_i, \omega_j) \Leftrightarrow j = \arg \min_{j=1, \dots, c} \mathbf{B}_G(f_{R_i}, f_{\omega_j}). \quad (2)$$

2.2. SVM e Classificação Baseada em Regiões

SVM é um método de classificação que tem se tornado cada vez mais popular nas pesquisas em sensoriamento remoto. Este método consiste em realizar a separação entre amostras de treinamento a partir de um hiperplano que possui maior margem de separação. Tal hiperplano

corresponde ao lugar geométrico onde a seguinte função é nula:

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b, \quad (3)$$

sendo \mathbf{w} o vetor ortogonal ao hiperplano e $|b|/\|\mathbf{w}\|$ a distância entre o hiperplano e a origem do espaço de atributos.

A partir de $\mathcal{D}_l = \{(\mathbf{x}_i, \omega_j) \in \mathcal{X} \times \Omega : i = 1, \dots, m; j = 1, 2\}$, e considerando inicialmente o conjunto de classes $\Omega = \{\omega_1, \omega_2\}$, os parâmetros \mathbf{w} e b de (3) são obtidos a partir da solução do seguinte problema de otimização (THEODORIDIS; KOUTROUMBAS, 2008):

$$\begin{aligned} & \max_{\lambda} \sum_{u=1}^m \lambda_u - \frac{1}{2} \sum_{u=1}^m \sum_{v=1}^m \lambda_u \lambda_v y_u y_v \langle \mathbf{x}_u, \mathbf{x}_v \rangle \\ & \text{sujeito a: } \begin{cases} 0 \leq \lambda_u \leq C, u = 1, \dots, m \\ \sum_{u=1}^m \lambda_u y_u = 0 \end{cases}, \end{aligned} \quad (4)$$

onde λ_u multiplicadores de Lagrange y_u o indicador da classe de \mathbf{x}_u , onde y_u é +1 se (\mathbf{x}_u, ω_1) ou -1 se (\mathbf{x}_u, ω_2) . O escalar C é um parâmetro (penalidade) que atua no processo de treinamento do método. Sendo $SV = \{\mathbf{x}_u \in \mathcal{X} : \lambda_u \neq 0; u = 1, \dots, n \leq m\}$ o conjunto de vetores suporte, são computados: $\mathbf{w} = \sum_{\mathbf{x}_u \in SV} \lambda_u y_u \mathbf{x}_u$ e $b = \frac{1}{\#SV} (\sum_{\mathbf{x}_u \in SV} y_u + \sum_{\mathbf{x}_u \in SV} \sum_{\mathbf{x}_v \in SV} \lambda_u \lambda_v y_u y_v \langle \mathbf{x}_u, \mathbf{x}_v \rangle)$.

Uma vez definido (3), o padrão \mathbf{x} é associado a ω_1 se $f(\mathbf{x}) \geq 0$, ou associado a ω_2 se $f(\mathbf{x}) < 0$, o que caracteriza o método SVM como um classificador binário. Por sua vez, a aplicação do SVM em problemas com mais de duas classes deve estar vinculada a adoção de uma estratégia multiclases, como por exemplo, a estratégia Um-Contra-Todos (THEODORIDIS; KOUTROUMBAS, 2008).

O produto interno $\langle \mathbf{x}_u, \mathbf{x}_v \rangle$ presente em (4), e por sua vez nos cálculos de \mathbf{w} e b , pode ser substituído por funções *kernel* $K : \mathcal{X}^2 \rightarrow \mathbb{R}$. Tais funções são usualmente adotadas para realização de classificação entre padrões não linearmente separáveis. Outro propósito destas funções é a generalização da aplicação do método SVM em problemas cujos padrões não possuem representação vetorial, como por exemplo, cadeias de caracteres e conjuntos de diferentes cardinalidades.

Esta flexibilidade oferecida pelas funções *kernel* possibilita uma forma distinta de aplicar SVM nos problemas de classificação baseada em regiões. Através do emprego da função *kernel* de Bhattacharyya (KONDOR; JEBARA, 2003) é possível considerar cada região da imagem (observada como um conjunto de pixels) como um padrão único. Tal função é definida por:

$$K(R_u, R_v) = e^{-\alpha \mathbf{B}_G(R_u, R_v)}, \quad (5)$$

onde R_u e R_v são regiões, \mathbf{B}_G é a distância de Bhattacharyya definida em (1) e $\alpha \in \mathbb{R}$ é um parâmetro ajustável.

2.3. Grafo e Classificação Baseada em Regiões

Dentre diferentes peculiaridades, os métodos de Classificação por Grafo são caracterizados pelo aprendizado semissupervisionado. Este tipo de aprendizado é motivado quando há escassez de dados rotulados suficientes para o devido treinamento do classificador. Nestas circunstâncias, o problema de insuficiência de amostras para um treinamento adequado pode ser minimizado com o aumento do conjunto de treinamento, utilizando padrões não rotulados, os quais são abundantes na maioria dos problemas de classificação (CIRELO; COZMAN, 2003).

De modo geral, a Classificação por Grafo emprega uma matriz de afinidade G , a qual é uma representação numérica de um grafo. Nesta matriz são representadas as similaridades entre

padrões, sejam rotulados ou não. Formalmente, sendo \mathcal{D}_l e \mathcal{D}_u respectivamente conjuntos de dados/padrões rotulados e não rotulados, $\mathbf{x}_r, \mathbf{x}_s \in \mathcal{D}_l \cup \mathcal{D}_u$ representam dois vértices no grafo, cujo valor (peso) associado a aresta entre tais vértices corresponde a medida de similaridade g_{rs} , a qual é um elemento de G . Os padrões \mathbf{x}_r e \mathbf{x}_s tendem a estar associados a mesma classe a medida que o valor de g_{rs} aumenta.

Segundo Zhu e Goldberg (2009), os métodos baseados em grafo são fundamentados na hipótese de “suavidade dos rótulos”, ou seja, a mudança da classe associada aos padrões, nesse caso os vértices, varia de forma suave sobre o grafo. O processo de associação de uma classe a um vértice (padrão) não rotulado depende da similaridade apresentada com relação aos demais vértices do grafo.

Dentre diferentes propostas apresentadas na literatura, em Camps-Valls, Tatyana e Zhou (2007) é apresentado um método de Classificação por Grafo que permite o uso de função *kernel*. Para isso, seja $\mathcal{D} = \mathcal{D}_l \cup \mathcal{D}_u$ um conjunto composto por m padrões rotulados e n não rotulados, a matriz de afinidade G pode ser determinada por:

$$G_{(m+n) \times (m+n)} : g_{rs} = e^{-\frac{\|\mathbf{x}_r - \mathbf{x}_s\|^2}{2\sigma^2}}; \quad r, s = 1, \dots, m+n; \sigma \in \mathbb{R}_+. \quad (6)$$

Por sua vez, verifica-se neste presente trabalho que $e^{-\frac{\|\mathbf{x}_r - \mathbf{x}_s\|^2}{2\sigma^2}}$ pode ser substituída pela função *kernel* Bhattacharyya, definida em (5), uma vez que ambas são exemplos diferentes de funções *kernel*. Sob esta condição, o método passa a tratar o problema de classificação como um caso baseado em regiões.

Em seguida, sobre os valores de similaridade de G é aplicada a seguinte normalização simétrica:

$$S = \sqrt{Q^{-1}}G\sqrt{Q^{-1}}, \quad (7)$$

onde Q é uma matriz diagonal, denominada Matriz Grau, tal que $q_{rr} = \sum_r g_{rs}$.

Com relação aos padrões de \mathcal{D} (agora considerados como regiões) é determinada a matriz de rótulos Y , definida por:

$$Y_{(m+n) \times c} : y_{rj} = \begin{cases} 1 & \text{se } (R_r, \omega_j) \\ 0 & \text{caso contrário} \end{cases} \quad r = 1, \dots, m+n; \quad j = 1, \dots, c, \quad (8)$$

Cabe ressaltar que Y possui o número de linhas equivalente ao número de padrões envolvidos no problema de classificação, enquanto o número de colunas refere-se a quantidade de classes do problema. Observa-se que as linhas de Y referentes aos padrões não rotulados são nulas. Por fim, a partir das matrizes S e Y é determinada a matriz U :

$$U = (I - \beta S)^{-1}Y, \quad (9)$$

sendo I é a matriz identidade e $\beta \in (0, 1)$ é um parâmetro de regularização.

Para classificação dos padrões não rotulados é utilizada a seguinte regra:

$$(\mathbf{x}_i, \omega_j) \Leftrightarrow \arg \max_{j=1, \dots, c} (u_{ij}), \quad (10)$$

onde u_{ij} é elemento de $U_{(m+n) \times c}$.

3. Experimentos

Nas seções anteriores foram discutidos três métodos para Classificação Baseada em Regiões, sendo eles: Classificador de Mínima Distância Estocástica (Seção 2.1), SVM (Seção 2.2) e Classificação por Grafo (Seção 2.3). Em especial, a aplicação dos dois últimos

métodos citados para Classificação Baseada em Regiões é suportada pelo uso da função *kernel* Bhattacharyya.

Nesta seção é apresentado um estudo de caso onde tais métodos são empregados na classificação do uso e cobertura do solo, tendo como área de estudo uma região compreendida pelo entorno da Floresta Nacional do Tapajós, no Estado do Pará. Para este estudo, foi utilizada uma imagem obtida pelo satélite SPOT, obtida da área de estudo em setembro de 2009. Tal imagem, ilustrada na Figura 1(a), possui dimensão de 2212×1423 pixels, resolução de aproximadamente 2,5 metros e três bandas espectrais referentes aos comprimentos de onda do vermelho, verde e azul. O método de segmentação por Crescimento de Regiões disponível no SIG SPRING (CAMARA et al., 1996) foi empregado na segmentação da imagem de estudo. A informação definida pela segmentação é empregada na definição dos limites das diferentes regiões que compõe a cena, sendo assim fundamental para os métodos de classificação empregados neste estudo.

Dentre os diferentes tipos de uso e cobertura do solo existentes na cena, foram identificadas áreas de Floresta Primária, dois tipos de Solo Exposto (limpo e em repouso), Pastagem não Manejada (i.e. pasto sujo), dois tipos de Pastagens Manejadas (com e sem presença de Palmas) e Sucessão Secundária Inicial. Estas informações foram obtidas a partir de um levantamento de campo realizado no mesmo mês e ano (setembro de 2009) da imagem empregada neste estudo. A Figura 1(b) apresenta a localização das amostras utilizadas no treinamento dos métodos e na avaliação dos resultados de classificação. A quantidade de polígonos e pixels que compreendem as amostras de treinamento e validação, referentes às classes consideradas, são apresentadas na Tabela 1.

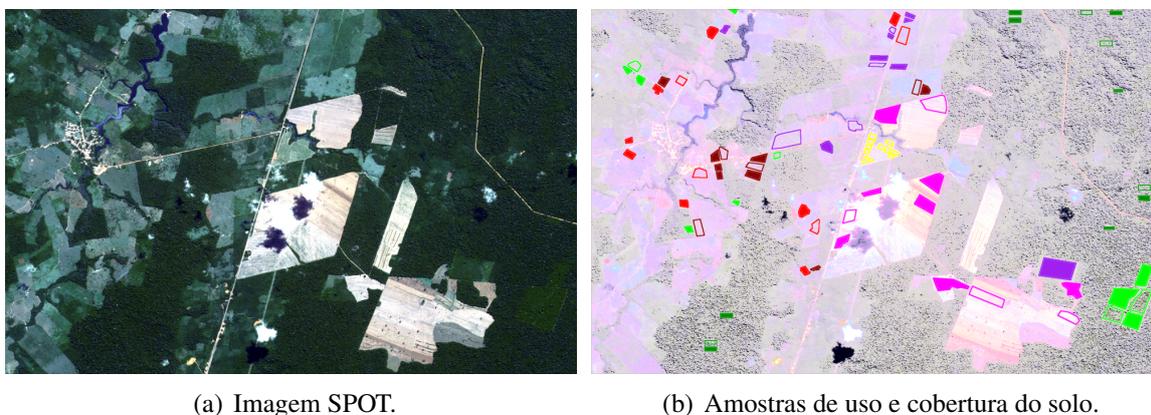


Figura 1: Imagens SPOT e amostras de uso e cobertura do solo empregadas no estudo de caso. Polígonos cheios representam amostras de treinamento e polígonos vazios as amostras de validação.

Uma vez definidas a área de estudo, a imagem utilizada no experimento e as amostras dos diferentes tipos de uso e cobertura do solo presentes na cena, foi realizado o processo de classificação através do Classificador de Mínima Distância Estocástica (CMDE), SVM e Classificação por Grafo (CG). O parâmetro α presente na função *kernel* (Equação 5), assim como os parâmetros penalidade (C) e regularizador (β) dos métodos SVM e CG, respectivamente, foram selecionados através de um procedimento exaustivo de Busca em Grade. Este procedimento indicou que para o método SVM a melhor configuração é $\alpha = 2,5$ e $C = 1000$, enquanto para CG a configuração adequada é $\alpha = 1,5$ e $\beta = 0,95$. Cabe ainda mencionar que para aplicação do SVM neste estudo, foi adotada a estratégia multiclases Um-Contra-Todos.

Tabela 1: Resumo sobre as amostras de uso e cobertura do solo utilizadas no estudo.

<i>Classes</i>	<i>Cor</i>	<i>Treinamento (pixels/pols.)</i>	<i>Teste (pixels/pols.)</i>
Pastagem Manejada	●	13797/8	2827/4
Solo Exposto Limpo	●	34772/7	11265/4
Pastagem Manejada com Palmas	●	12983/8	2584/4
Solo Exposto em Repouso	●	3523/7	1033/4
Sucessão Secundária Inicial	●	24117/8	4094/4
Pastagem Não Manejada	●	22783/8	2850/4
Floresta Primária	●	8633/8	1914/4

Os resultados obtidos por cada um dos métodos são apresentados na Figura 2. A partir das amostras de teste identificadas na Figura 1(b), o Coeficiente de Concordância *Kappa* (κ) (CONGALTON; GREEN, 2009) foi empregado para mensurar a acurácia dos resultados de classificação. Os valores de acurácia registrados a partir deste coeficiente foram: 0,67, com desvio padrão de 0,011, para método CMDE; 0,7, com desvio padrão de 0,011 para o método SVM; e 0,77, com desvio padrão de 0,013, para o método CG.

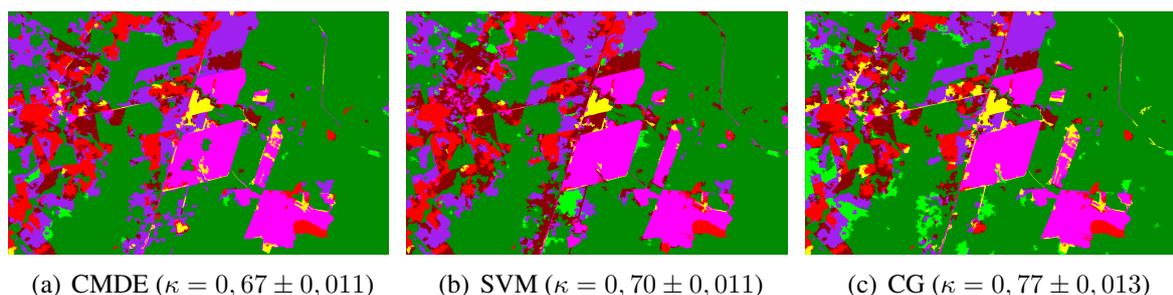


Figura 2: Resultados de classificação obtidos pelos diferentes métodos de classificação adotados.

Os resultados obtidos mostram que os métodos SVM e CG alcançaram desempenhos superiores comparados ao CMDE. Baseado em um teste de hipótese bilateral para comparação entre os valores de coeficientes *Kappa* (CONGALTON; GREEN, 2009), verifica-se com 95% de confiança que o valor de acurácia atingido por CG é estatisticamente superior aos demais métodos.

Ao comparar os resultados de classificação ilustrados na Figura 2, é possível observar que o método CG foi mais acurado na classificação das áreas de Sucessão Secundária Inicial, assim como uma melhor delimitação entre as áreas de pastagem e Floresta Primária.

4. Conclusões e Perspectivas Futuras

Este trabalho consistiu em verificar a aplicação da Classificação por Grafo, associada ao uso de distância estocástica, em um problema de classificação baseada em regiões. Como referencial, foram adotados métodos já empregados anteriormente nesta abordagem de classificação e que também utilizam o conceito de distância estocástica. Os resultados obtidos no estudo de caso realizado evidenciaram potencial no método investigado.

Os resultados subsidiam concluir que a forma de aprendizado do método CG (semisupervisionado) assim como a hipótese de suavidade dos “rótulos” influenciaram na obtenção de resultados mais acurados.

Como perspectivas futuras, é necessário verificar o desempenho da Classificação por Grafo em outros problemas de classificação baseada em regiões, envolvendo imagens de outros sensores ou até mesmo através de imagens simuladas. É possível ainda investigar a utilização de outras distâncias estocásticas e/ou funções *kernel* por este método.

Agradecimentos

Os autores agradecem à FAPESP (proc. 2014/14830-8), CNPQ (proc. 151571/2013-9) e CAPES pelo auxílio financeiro.

Referências

- CAMARA, G. et al. Spring: Integrating remote sensing and gis by object-oriented data modelling. *Computers & Graphics*, v. 20, p. 3, 1996.
- CAMPS-VALLS, G.; TATYANA, V. B.; ZHOU, D. Semi-supervised graph-based hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, v. 45, p. 2044–3054, 2007.
- CIRELO, M. C.; COZMAN, F. G. Aprendizado de semi-supervisionado de classificadores bayesianos utilizando testes de independência. In: IV ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL. Campinas, 2003.
- CONGALTON, R. G.; GREEN, K. *Assessing the accuracy of remotely sensed data: principles and practices*. Boca Raton: CRC Press, 2009. 183 p. ISBN 978-1-4200-5512-2.
- DESHENG, L.; FAN, X. Assessing object-based classification: advantages and limitations. *Remote Sensing Letters*, Taylor & Francis, v. 1, n. 4, p. 187–194, 2010.
- GIGANDET, X. et al. Region-based satellite image classification: method and validation. In: *IEEE International Conference on Image Processing*. [S.l.: s.n.], 2005. v. 3, p. 832–835.
- KONDOR, R.; JEBARA, T. A kernel between sets of vectors. In: *In International Conference on Machine Learning*. [S.l.: s.n.], 2003.
- NEGRI, R. G.; DUTRA, L. V.; SANT'ANNA, S. J. S. Support Vector Machine and Bhattacharyya kernel function for regionbased classification. In: INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 2012, Munich. *Proceedings...* Munich: IEEE, 2012. p. 5422–5425.
- RICHARDS, J. A.; XIUPING, J. *Remote Sensing Digital Image Analysis: An Introduction*. New York: Springer, 2005. ISBN 3540251286.
- SILVA, W. B. et al. Land cover discrimination at brazilian amazon using region based classifier and stochastic distance. In: *2011 IEEE International Geoscience and Remote Sensing Symposium*. [S.l.: s.n.], 2011. p. 2900–2903.
- THEODORIDIS, S.; KOUTROUMBAS, K. *Pattern Recognition*. 4th. ed. San Diego: Academic Press, 2008. 984 p. ISBN 9781597492720.
- ZHU, X.; GOLDBERG, A. B. *Introduction to Semi-Supervised Learning*. [S.l.]: Morgan & Claypool Publishers, 2009.