

Desenvolvimento de Novas Funções Kernel para Classificação Contextual de Imagens

Rogério Galante Negri

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@ict.unesp.br

Abstract. Kernel functions have revolutionized the theory and practice in Pattern Recognition, and consequently the image classification applications. Besides allows the definition of non-linear versions of methods like Support Vector Machine (SVM), such functions allow generalize the application of these methods on the classification of non-vector patterns, such as probability distributions, information sets, etc. This possibility motivates the development of kernel functions able to deal with the context which the pixels are inserted and consequently inducing contextual classifications when adopted. In this initial study, three kernel functions are proposed for contextual classification. These functions are based on stochastic distances, non-parametric statistical tests and spatial variation modeling. A case study about the land use and land cover classification with an ALOS PALSAR image is carried in order to compare the performance of the SVM method though the use of the developed kernel functions. Comparisons with other contextual methods based on SVM are included in this analyzes. The results shows potential on the new proposals, especially the kernels based on stochastic distance and nonparametric statistical test.

Keywords: kernel function, context, image classification funções kernel, contexto, classificação de imagens

1. Introdução

A utilização de *funções kernel* provocou uma verdadeira revolução na teoria e prática do Reconhecimento de Padrões, e conseqüentemente nas aplicações de classificação de imagens, por permitir criar versões não lineares dos métodos clássicos lineares, dentre os quais está incluso a Máquina de Vetores Suporte (*Support Vector Machine* - SVM), um dos mais versáteis métodos já propostos. Recentemente, as funções kernel permitiram ainda generalizar tais métodos para padrões não vetoriais, como por exemplo, conjuntos, sequências de símbolos e grafos. Assim, pode-se afirmar que as funções kernel permitem adaptar o método às características do problema (dados), e não o contrário (SCHÖLKOPF; SMOLA, 2002).

A fim de aumentar a acurácia dos resultados de classificação, a inclusão da informação contextual tem sido uma linha de investigação quanto o desenvolvimento de métodos de classificação de imagens. A investigação de funções kernel capazes de incluir a informação contextual e por sua vez auxiliar na obtenção de resultados mais precisos através de métodos baseado em kernel não é uma exceção, como por exemplo, no estudo apresentado em Gurrum e Kwon (2013).

Neste trabalho são apresentas três propostas de funções kernel para classificação contextual de imagens de sensoriamento remoto. Tais propostas são baseadas em conceitos derivados de distâncias estocásticas, testes de hipóteses não paramétricos e em uma modelagem da variabilidade espacial dos pixels.

Um estudo de caso sobre a classificação do uso e cobertura do solo, fazendo uso de uma imagem proveniente do sensor PALSAR, abordo do satélite ALOS, envolvendo o emprego das propostas formalizadas neste trabalho, além de outras propostas já presentes na literatura, é apresentado.

2. Classificação Contextual de Imagens e Funções Kernel

Um classificador é representado por uma função $F : \mathcal{X} \rightarrow \mathcal{Y}$, que associa elementos do espaço de atributos \mathcal{X} a uma das classes de $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$, com $c \in \mathbb{N}^*$, a partir de um dado rótulo (indicador) de classe em $\mathcal{Y} = \{1, 2, \dots, c\}$. Nestas condições, para $\mathbf{x} \in \mathcal{X}$ e $y \in \mathcal{Y}$, $y = F(\mathbf{x})$ indica que \mathbf{x} pertence à classe ω_y .

Admitindo a abordagem pontual, a classificação de imagens consiste na aplicação de F sobre os padrões (vetor de atributos de um pixel s) que compõem uma imagem \mathcal{I} , definida sobre um reticulado (suporte) $\mathcal{S} \subset \mathbb{N}^2$, cujo resultado de classificação pode ser denotado por $F(\mathcal{I})$. Com relação à imagem em que é conduzido o processo de classificação, $\mathcal{I}(s) = \mathbf{x}$ denota que o pixel $s \in \mathcal{S}$ de \mathcal{I} possui atributos representados por $\mathbf{x} \in \mathcal{X}$. Ainda, as posições ocupadas pelos vizinhos de s pertencem ao conjunto $\mathcal{V}_\rho(s) = \{t \in \mathcal{S} : 0 \leq md(s, t) \leq \rho\}$, sendo ρ denominado *raio de influência da vizinhança* e $md(\cdot, \cdot)$ é uma distância adotada. Nestas condições, define-se por contexto de s o conjunto de pares ordenados (\mathbf{x}, y) , compostos pelo vetor de atributos \mathbf{x}_i e o indicador de classe y_i , tal que $\mathcal{I}(t) = \mathbf{x}_i$ e $t \in \mathcal{V}_\rho(s)$.

Os diferentes métodos de classificação de imagem propostos na literatura podem ser entendidos como maneiras distintas de modelar a função $F : \mathcal{X} \rightarrow \mathcal{Y}$ e aplicá-la na classificação de \mathcal{I} . Para os métodos cujo aprendizado é supervisionado, são utilizadas informações disponíveis em um *conjunto de treinamento* $\mathcal{D} = \{(\mathbf{x}_i, y_i) \in \mathcal{X} \times \mathcal{Y} : i = 1, \dots, m\}$ com $m \in \mathbb{N}^*$. O mapeamento entre \mathcal{X} e \mathcal{Y} definido por F representa o conhecimento adquirido das informações observadas em \mathcal{D} .

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:

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b, \quad (1)$$

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 determinação de \mathbf{w} e b é realizada mediante a resolução de um problema de otimização construído em função dos elementos definidos em \mathcal{D} . Maiores detalhes a respeito do processo de treinamento do método SVM podem ser encontrados em Webb (2002).

A Equação 1 consiste em um hiperplano de separação definido no espaço de atributo original dos dados. O mapeamento dos padrões envolvidos no processo de classificação, entre os espaço de atributos original e um novo espaço de atributos é uma estratégia usualmente empregada para aumentar a acurácia do método SVM. Tal mapeamento e classificação em outro espaço de atributos pode ser realizado de forma implícita através do emprego de *funções kernel*, as quais substituem o produto interno presente nesta equação e no problema de otimização que consiste o treinamento do método. Outro propósito destas funções é a generalização da aplicação do método SVM, ou qualquer método baseado em kernel, em problemas cujos padrões não possuem representação vetorial, como por exemplo, cadeias de caracteres e conjuntos de diferentes cardinalidades.

Formalmente, $K : \mathcal{X}^2 \rightarrow \mathbb{R}$ é caracterizada como uma função kernel se é simétrica com relação aos padrões de entrada e atender as condições do Teorema de Mercer (THEODORIDIS; KOUTROUMBAS, 2008). A definição de uma função kernel e respectiva verificação do teorema mencionado pode não ser uma tarefa trivial. Uma maneira de definir funções kernel é através da adoção de modelos gerais apresentados na literatura, como por exemplo, as funções de base radial (SCHÖLKOPF; SMOLA, 2002):

$$K(\mathbf{x}_u, \mathbf{x}_v) = g(d(\mathbf{x}_u, \mathbf{x}_v)), \quad (2)$$

onde $g : \mathbb{R} \rightarrow \mathbb{R}$ é uma função real estritamente positiva e $d : \mathcal{X}^2 \rightarrow \mathbb{R}$ é uma métrica.

A proposta de novas funções kernel para classificação de imagens de sensoriamento remoto tem sido um tema de investigação relevante. O conceito da indução de classificações contextuais a partir de funções kernel é apresentado pela primeira vez em Camps-Valls et al. (2006) e posteriormente por Gurram e Kwon (2013). Em ambas as pesquisas, a informação espacial, delimitada por um raio de influência, é integrada na forma de uma função kernel média. Os resultados produzidos tendem a assumir um comportamento homogêneo, semelhante a uma suavização pelo filtro de média. Outros exemplos de propostas de classificação contextual via funções kernel são verificadas em Liu et al. (2013) e Huo et al. (2015). Partindo do estudo anteriormente apresentada por Kondor e Jebara (2003), foi proposto em Negri et al. (2016) o uso da função kernel Bathacharryya para lidar com a abordagem de classificação baseada em regiões.

Considerando o modelo de função kernel de base radial, mencionado anteriormente, verifica-se a possibilidade desenvolver funções capazes de integrar a informação do contexto espacial dos padrões/pixels a fim de realizar classificação contextual de imagens. Nas subseções seguintes são apresentadas três funções kernel capazes de incluir a informação contextual no processo de classificação em métodos baseados em kernel, como por exemplo, o SVM.

2.1. Função Kernel Jeffries-Matusita

Uma maneira direta de considerar a informação da vizinhança em um processo de classificação contextual é admitir tal vizinhança como um único padrão. Desta forma, os padrões passam a não ser mais representados por vetores $\mathbf{x} \in \mathcal{X}$, como definidos na Seção 2, mas sim pela própria vizinhança $\mathcal{V}_\rho(s)$. Sob consideração do modelo apresentado na Equação 2, a escolha de $d(\cdot, \cdot)$ como uma distâncias estocásticas torna-se um meio conveniente para comparação entre diferentes padrões (ou conjuntos de padrões) $\mathcal{V}_\rho(s_i)$ e $\mathcal{V}_\rho(s_j)$.

Este tipo de distância quantifica a separabilidade entre dois conjuntos de informação mediante as dissimilaridades entre as distribuições de probabilidade que modelam a informação de tais conjuntos. A distância Jeffries-Matusita (JM) é usualmente empregada nas aplicações envolvendo processamento de imagens de sensoriamento remoto. Considerando que as informações em $\mathcal{V}_\rho(s_i)$ e $\mathcal{V}_\rho(s_j)$ podem ser modeladas segundo a distribuição Gaussiana Multivariada, a distância JM é dada por (RICHARDS; XIUPING, 2005):

$$JM(\mathcal{V}_\rho(s_i), \mathcal{V}_\rho(s_j)) = 2 \left(1 - e^{-B(\mathcal{V}_\rho(s_i), \mathcal{V}_\rho(s_j))} \right), [0, 2]; \quad (3)$$

onde $B(\cdot, \cdot)$ é a distância de Bhattacharyya assumindo distribuição Gaussiana Multivariada dos dados, expressa por:

$$B(\mathcal{V}_\rho(s_i), \mathcal{V}_\rho(s_j)) = \frac{1}{8} (\mu_{\mathcal{V}_\rho(s_i)} - \mu_{\mathcal{V}_\rho(s_j)})^T \left(\frac{\Sigma_{\mathcal{V}_\rho(s_i)} + \Sigma_{\mathcal{V}_\rho(s_j)}}{2} \right)^{-1} (\mu_{\mathcal{V}_\rho(s_i)} - \mu_{\mathcal{V}_\rho(s_j)}) + \frac{1}{2} \ln \left(\frac{|\Sigma_{\mathcal{V}_\rho(s_i)} + \Sigma_{\mathcal{V}_\rho(s_j)}|}{\sqrt{|\Sigma_{\mathcal{V}_\rho(s_i)}| |\Sigma_{\mathcal{V}_\rho(s_j)}|}} \right), \quad (4)$$

onde $\mu_{\mathcal{V}_\rho(s_z)}$ e $\Sigma_{\mathcal{V}_\rho(s_z)}$ são o vetor média e a matriz de covariância estimadas a partir de $\mathcal{V}_\rho(s_z)$.

A substituição direta de $d(\cdot, \cdot)$ presente na Equação 2 pela expressão distância JM não proporciona uma função kernel válida, uma vez que tal distância estocástica não é uma métrica por não atender a propriedade da desigualdade triangular. No entanto, este problema pode ser contornado por meio da adição de uma constante suficientemente grande à expressão original da distância quando os padrões $\mathcal{V}_\rho(s_i)$ e $\mathcal{V}_\rho(s_j)$ são diferentes. Uma vez que a imagem de função JM está limitada ao intervalo $[0, 2]$, a adição de 2 na Equação 3 irá proporcionar uma métrica.

Após as considerações apresentadas, é definida a *função kernel Jeffries-Matusita*:

$$K_{JM}(\mathcal{V}(s_i), \mathcal{V}(s_j)) = e^{-\gamma(JM(\mathcal{V}_\rho(s_i), \mathcal{V}_\rho(s_j))+2)}, \quad (5)$$

onde $\gamma \in \mathbb{R}_+$ é um parâmetro de regularização.

2.2. Função Kernel Kruskal-Wallis

Seguindo considerações análogas às apresentadas na Seção 2.1 a respeito da vizinhança como um único padrão e do uso do modelo de base radial para definição de funções kernel, uma alternativa para comparação das informações presentes em $\mathcal{V}_\rho(s_i)$ e $\mathcal{V}_\rho(s_j)$ consiste na aplicação de teste de hipóteses, seguido pela observação do p -valor fornecido por tal teste.

Dentre diferentes técnicas presentes na literatura para comparação entre conjuntos de dados, o teste paramétrico t é usualmente empregado. No entanto, tal teste pode proporcionar resultados inadequados quando o conjunto de dados, em função de sua dimensão ou mesmo comportamento, não atende a hipótese de normalidade. Em casos como este, o uso de testes não paramétricos é uma alternativa, como por exemplo, o teste de Kruskal-Wallis (MOOD; BOES; GRAYBILL, 1974).

Considerando, em primeiro momento, $p_{kw}(\mathcal{V}_\rho(s_i), \mathcal{V}_\rho(s_j); b)$ como uma função que retorna o p -valor fornecido pelo teste de Kruskal-Wallis ao compara as observações (unidimensionais) inseridas nas vizinhanças de s_i e s_j com relação apenas à b -ésima banda da imagem \mathcal{I} , é definida a seguinte medida de similaridade:

$$P(\mathcal{V}_\rho(s_i), \mathcal{V}_\rho(s_j)) = \frac{1}{b} \sum_{k=1}^b (1 - p_{kw}(\mathcal{V}_\rho(s_i), \mathcal{V}_\rho(s_j); k)), \quad [0, 1]. \quad (6)$$

Cabe observar que ao passo que a similaridade de entre $\mathcal{V}_\rho(s_i)$ e $\mathcal{V}_\rho(s_j)$ aumenta, o p -valor resultante tende a 1. A fim de proporcionar uma relação direta entre similaridade/distância e p -valor, onde padrões idênticos apresentam distância nula (i.e., a propriedade de identidade), foi feita a adoção de $1 - p_{kw}(\cdot, \cdot; \cdot)$ na Equação 6. Uma vez que a $P(\cdot, \cdot)$ é simétrica, positivamente definida e atende a propriedade de identidade, tal função é considerada uma medida de distância. Adotando o mesmo procedimento apresentado na Seção 2.1, é possível definir uma métrica a partir de $P(\cdot, \cdot)$ através da adição da constante 1 quando $\mathcal{V}_\rho(s_i) \neq \mathcal{V}_\rho(s_j)$.

Novamente, a partir do modelo de função kernel de base radial, é definida a *função kernel Kruskal-Wallis*:

$$K_{KW}(\mathcal{V}(s_i), \mathcal{V}(s_j)) = e^{-\gamma(2-P(\mathcal{V}_\rho(s_i), \mathcal{V}_\rho(s_j)))} \quad (7)$$

onde $\gamma \in \mathbb{R}_+$, assim como na Equação 5, é um parâmetro de regularização.

2.3. Função Kernel do Espaço Estendido

A terceira função kernel apresentada, diferente das propostas introduzidas nas Seções 2.1 e 2.2, baseadas em comparações envolvendo todas as informações das respectivas vizinhanças e de modo simultâneo, visa comparar os elementos centrais de cada vizinhança somadas as respectivas flutuações de variação nos diferentes valores de alcance nesta vizinhança, ponderadas pelo afastamento do elemento central.

Para a vizinhança $\mathcal{V}_\rho(s_i)$, o comportamento do elemento central s_i somado às flutuações de variação em um dado alcance r é determinado por:

$$F(s_i, r) = \frac{1}{\#\mathcal{V}_{\rho=r}(s_i)} \sum_{s_k \in \mathcal{V}_{\rho=r}(s_i)} (\mathbf{x}_i - \mathbf{x}_k), \quad (8)$$

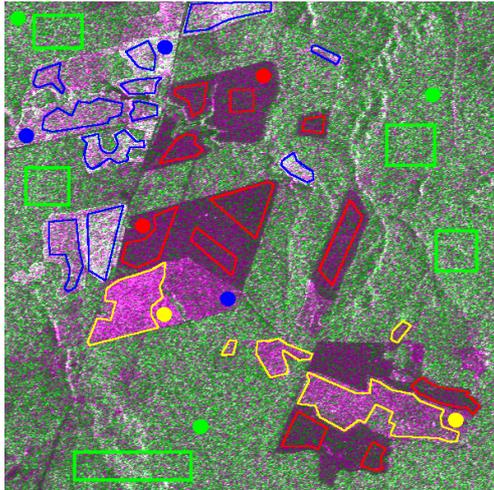


Figura 1: Imagem ALOS-PALSAR e amostras usadas para treinamento e teste.

Tabela 1: Amostras de uso e cobertura do solo localizados na área de estudo.

<i>Classe</i>	<i>Cor</i>	<i>Treino/Teste</i>
Floresta	●	768/14136
Pastagem	●	768/14678
Solo Exposto	●	512/14207
Agricultura	●	512/14395

onde $\mathcal{V}_{\rho=r}(s_i)$ representa o subconjunto de elementos em $\mathcal{V}_{\rho=r}(s_i)$ cuja distância do elemento central s_i é r .

Uma vez definida a expressão que modela as variações do contexto em função da distância ao elemento central, é definida a *função kernel do Espaço Estendido*:

$$K_{EE}(\mathcal{V}(s_i), \mathcal{V}(s_j)) = \sum_{r=0}^{\rho} (e^{-\gamma \|F(s_i,r) - F(s_j,r)\|} \times e^{-\eta \rho}). \quad (9)$$

onde $\gamma \in \mathbb{R}_+$ é responsável por regularizar a informação da variação espacial e $\eta \in \mathbb{R}_+$ pondera a influência da variação espacial na caracterização do contexto em função da distância com que os elementos da vizinhança estão distribuídos.

3. Experimentos e Resultados

Esta seção apresenta um estudo de caso sobre o emprego e comparação das funções kernel definidas nas Seções 2.1, 2.2 e 2.3 pelo método SVM na classificação do uso e cobertura do solo em uma área de estudo compreendida pelo entorno da Floresta Nacional do Tapajós, no Estado do Pará. Uma imagem obtida em setembro de 2009, pelo sensor PALSAR abordo do satélite ALOS, foi empregado neste estudo. Tal imagem possui 20 m de resolução, dimensão de 600×600 pixels e informação dos alvos segundo as polarizações HH e HV em amplitude.

Dentre os diferentes tipos de uso e cobertura do solo existentes na cena, foram considerados apenas áreas de Floresta Primária, Pastagem, Solo Exposto e Agricultura. Estas informações foram obtidas a partir de um levantamento de campo realizado no mesmo mês e ano da imagem empregada neste estudo. A Figura 1 ilustra a imagem empregada neste estudo, juntamente com a distribuição espacial das amostras dos tipos de uso e cobertura do solo considerados no estudo, distinguidas entre treinamento (polígonos sólidos) dos classificadores e teste/avaliação (polígonos vazios) dos resultados. A quantidade de pixels que compreendem as amostras de treinamento e teste, referentes às classes consideradas, são apresentadas na Tabela 1.

Comparações com outras versões contextuais do método SVM são incluídas neste estudo, a citar, a integração entre SVM e o algoritmo ICM (*Iterated Conditional Modes*) a suavização dos resultados de classificação gerados pelo SVM por meio da aplicação do filtro da moda, conforme discutido em Negri, Dutra e Sant'Anna (2016). Tais versões contextuais são denotadas por SVM+ICM e SVM+Moda.

Para caracterização da vizinhança, foi feito o emprego da distância do máximo¹ e dos valores 1, 2 e 3 como raio de influência de vizinhança (ρ), os quais definem por sua vez janelas espaciais de dimensão 3×3 , 5×5 e 7×7 , respectivamente. Tais vizinhanças foram consideradas pelo método SVM+Moda e pelo SVM com uso das funções kernel definidas nas Seções 2.1, 2.2 e 2.3.

Com relação aos métodos SVM, SVM+ICM e SVM+Moda, cabe mencionar que foi feito o emprego da função kernel RBF usual. Para lidar com a classificação de múltiplas classes (i.e., as quatro classes consideradas no estudo de caso), foi utilizada a estratégia Um-Contra-Todos (THEODORIDIS; KOUTROUMBAS, 2008).

A fim de mensurar a acurácia dos resultados de classificação proporcionados pelas funções kernel propostas neste estudo e pelos demais métodos considerados, foi feito uso do Coeficiente de Concordância *Kappa* (κ) e seu desvio padrão (CONGALTON; GREEN, 2009). A acurácia dos diferentes resultados de classificação foi computada com base nas amostras de teste ilustradas na Figura 1.

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 com uso do SVM e das diferentes funções kernel mencionadas anteriormente, além das versões contextuais SVM+ICM e SVM+Moda. Os parâmetros existentes nos diferentes kernel considerados foram ajustados através de um procedimento exaustivo de Busca em Grade. O gráfico apresentado na Figura 2 apresenta o desempenho do método SVM diante o uso das diferentes funções Kernel propostas e raios de influência da vizinhança (janelas espaciais), além dos métodos SVM, SVM+ICM e SVM+Moda. Os resultados de maior acurácia obtidos por cada uma das funções kernel e demais versões contextuais são apresentados na Figura 3.

Os índices de acurácia obtidos mostram que o uso das diferentes funções kernel proporcionam melhoras significativas em comparação o SVM usual, podendo ainda ser superior aos métodos SVM+ICM e SVM+Moda quando feito o uso dos kernels Kruskal-Wallis e Jefries-Matusita e janelas espaciais de dimensão conveniente. Ao observar os resultados de classificação ilustrados na Figura 3, é possível notar que, apesar do método SVM+Moda(7×7) estar relacionado a um dos maiores valores de acurácia obtido neste estudo de caso, a discriminação da classe Pastagem é inferior em comparação ao SVM+KJM(7×7). Ainda, dentre as três funções propostas, o Kernel do Espaço Estendido está associado aos mais baixos valores acurácia.

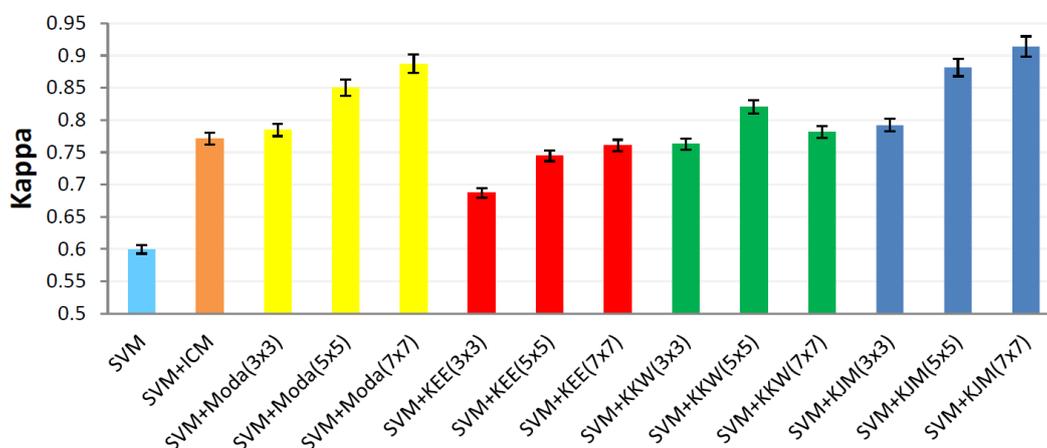


Figura 2: Desempenho dos métodos considerados no estudo de caso.

¹Sejam $s, t \in \mathbb{R}^2$ tal que $s = (s_1, s_2)$ e $t = (t_1, t_2)$, é denominada por *distância do máximo* a medida $md(s, t) = \max \{|s_1 - t_1|, |s_2 - t_2|\}$.

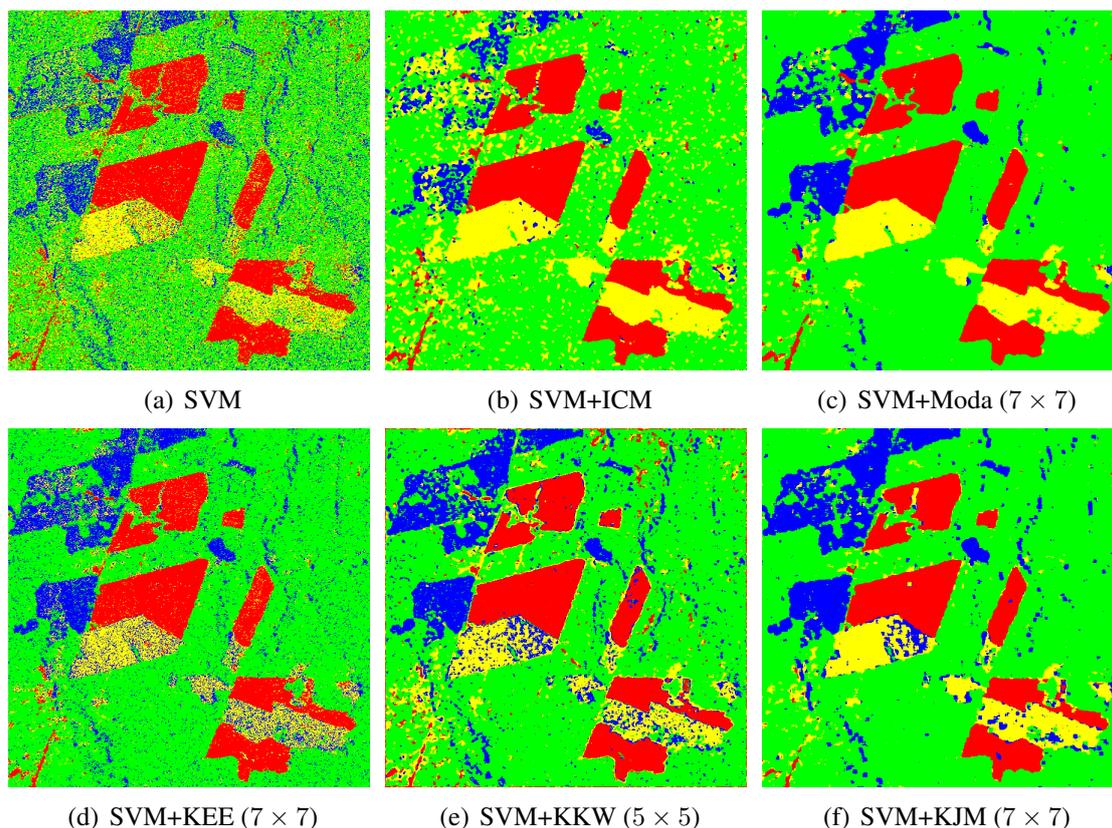


Figura 3: Resultados de classificação, segundo o coeficiente kappa, obtidos no estudo de caso. As barras de erro mensuram o desvio padrão do respectivo kappa calculado.

4. Conclusões e Perspectivas Futuras

De modo diferente de abordagens usuais, este trabalho visou a indução de resultados de classificação contextual pelo método SVM através do emprego de funções kernel. Para este fim, foram desenvolvidas três kernels diferentes, os quais são fundamentados em conceitos distintos. Tais conceitos referem-se a distâncias estocásticas, testes estatísticos não paramétricos e modelagem da variação espacial. Como referencial, foram adotados métodos contextuais baseados em SVM já discutidos na literatura.

Os resultados apresentados permitiram observar que as funções propostas, principalmente às baseadas em distâncias estocásticas e testes estatísticos não paramétricos, possuem potencial quanto a geração de resultados acurados. Em determinados casos, as funções propostas alcançam resultados mais precisos em comparação às versões contextuais consideradas no estudo.

Como perspectivas futuras a este estudo inicial, são citadas: a análise sobre a precisão de classificação nas bordas dos alvos; a consideração de outras estatísticas paramétricas e não paramétricas para o desenvolvimento de kernels segundo o conceito apresentado na Seção 2.2; e investigações mais profundas a respeito da modelagem da variação espacial partindo da motivação dada na Seção 2.3.

Agradecimentos

O autor agradece à FAPESP (proc. 2014/14830-8) pelo auxílio financeiro.

Referências

- CAMPS-VALLS, G. et al. Composite kernels for hyperspectral image classification. *Geoscience and Remote Sensing Letters, IEEE*, v. 3, n. 1, p. 93–97, 2006.
- 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.
- GURRAM, P.; KWON, H. Contextual SVM using Hilbert space embedding for hyperspectral classification. *IEEE Geoscience and Remote Sensing Letters*, v. 10, n. 5, p. 1031–1034, 2013.
- HUO, L. Z. et al. Semisupervised classification of remote sensing images with hierarchical spatial similarity. *IEEE Geoscience and Remote Sensing Letters*, v. 12, n. 1, p. 150–154, 2015. ISSN 1545-598X.
- KONDOR, R.; JEBARA, T. A kernel between sets of vectors. In: *In International Conference on Machine Learning*. [S.l.: s.n.], 2003.
- LIU, J. et al. Spatial-spectral kernel sparse representation for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 6, n. 6, p. 2462–2471, 2013. ISSN 1939-1404.
- MOOD, A. M.; BOES, D. C.; GRAYBILL, F. A. *Introduction to the theory of statistics*. 3th. ed. [S.l.]: McGraw-Hill, 1974. 564 p. ISBN 0070854653.
- NEGRI, R. G.; DUTRA, L. V.; SANT'ANNA, S. J. S. Comparing support vector machine contextual approaches for urban area classification. *Remote Sensing Letters*, v. 7, n. 5, p. 485–494, 2016.
- NEGRI, R. G. et al. Examining region-based methods for land cover classification using stochastic distances. *International Journal of Remote Sensing*, v. 37, n. 8, p. 1902–1921, 2016. Disponível em: <<http://dx.doi.org/10.1080/01431161.2016.1165883>>.
- RICHARDS, J. A.; XIUPING, J. *Remote Sensing Digital Image Analysis: An Introduction*. [S.l.]: Springer, 2005. ISBN 3540251286.
- SCHÖLKOPF, B.; SMOLA, A. J. *Learning with kernels : support vector machines, regularization, optimization, and beyond*. [S.l.]: MIT Press, 2002. (Adaptive computation and machine learning).
- THEODORIDIS, S.; KOUTROUMBAS, K. *Pattern Recognition*. 4th. ed. San Diego: Academic Press, 2008. 984 p. ISBN 9781597492720.
- WEBB, A. R. *Statistical Pattern Recognition*. 2nd. ed. Chichester: John Wiley & Sons, 2002. 514 p.