

## Uma Comparação entre Classificadores usando Regiões e Perfis EVI para Agricultura

Alexandre Noma<sup>1</sup>  
Thales Sehn Korting<sup>2</sup>  
Leila Maria Garcia Fonseca<sup>2</sup>

<sup>1</sup> Universidade Federal do ABC – UFABC  
Rua Santa Adélia, 166 – 09210-170 – Santo André – SP, Brasil  
alexandre.noma@ufabc.edu.br

<sup>2</sup> Instituto Nacional de Pesquisas Espaciais – INPE  
Caixa Postal 515 – 12245-970 – São José dos Campos - SP, Brasil  
thales, leila@dpi.inpe.br

**Abstract.** This work is within the Research Project entitled “*Land use Change in Amazônia: Institutional analysis and modelling at multiple temporal and spatial scales*”. The main goal is to develop techniques to help identifying and characterizing occupation patterns and trajectories in Amazonia, focusing on multitemporal data analysis. Computational tools, such as GeoDMA, have been developed to facilitate this kind of analysis. To enhance their capabilities, this paper compares different pattern recognition algorithms, ranging from classic ones, e.g. Naive Bayes and Nearest-Neighbor, to more popular and recent ones, e.g. Support Vector Machine and Optimum Path Forest. EVI profiles, acquired by MODIS sensor, were used as a multitemporal information to characterize two classes, agriculture and non-agriculture, in Mato Grosso state. All considered algorithms followed a region-based strategy, which can be considerably robust to noise and spectral variation. In this context, on one hand, regions were computed by using Baatz segmentation algorithm, which is a common approach for this purpose according to the literature. On the other hand, the segmentation algorithm was applied directly to the EVI profiles, which seems to be unusual but an important contribution, as the experimental results suggest. The classification algorithms were tested by using EVI profiles from June 2007 to June 2008 to match the currently annotated dataset from TerraClass Project. Finally, a comparison based on  $k$ -fold cross validation is provided as a preliminary benchmark to help choosing appropriate algorithms, for instance, to extend the GeoDMA software package.

**Keywords:** remote sensing, image processing, pattern recognition, sensoriamento remoto, processamento de imagens, reconhecimento de padrões

### 1. Introdução

Desde 1988, o Instituto Nacional de Pesquisas Espaciais (INPE) é o principal responsável pelo monitoramento da Floresta Amazônica. Dados provenientes de imagens de satélite permitem modelar o processo de desmatamento em larga escala para uma melhor compreensão sobre a dinâmica da paisagem. Neste contexto, dois projetos se destacam. O projeto PRODES (<http://www.obt.inpe.br/prodes/>) fornece mapas anuais de desmatamento na região Amazônica. Baseado nestes mapas, o projeto TerraClass ([http://www.inpe.br/cra/projetos\\_pesquisas/terraclass.php](http://www.inpe.br/cra/projetos_pesquisas/terraclass.php)) fornece uma classificação de uso e cobertura de solo. O objetivo é identificar diferentes tipos de uso nas regiões desmatadas, e.g. pastagem, agricultura, áreas urbanas, etc. Desta maneira, é possível não apenas fiscalizar, mas também desenvolver políticas públicas de prevenção para conter o avanço do desmatamento.

Dada a importância deste monitoramento, outros projetos estão atualmente em desenvolvimento. Por exemplo, o Projeto Temático “*Land use Change in Amazônia:*

*Institutional analysis and modelling at multiple temporal and spatial scales*” almeja desenvolver metodologias para identificar e caracterizar padrões e trajetórias de ocupação na região Amazônica. Inserido neste projeto, este trabalho explora dados multitemporais para identificar o uso dentro dos polígonos de desmatamento mapeados pelo PRODES, baseado nos temas definidos pelo TerraClass. Em particular, os dados temporais usados nos experimentos correspondem a uma sequência de imagens do sensor MODIS, adquiridas em diferentes instantes no tempo, e.g. perfis EVI com índices de vegetação. Informações temporais são fundamentais para determinar trajetórias de mudanças de uso e cobertura do solo, e.g. desmatamento, crescimento de culturas e modificação na estrutura das cidades. A ideia central é comparar diferentes técnicas de classificação para desenvolver e estender ferramentas para facilitar uma análise dos padrões de mudanças ocorridas nas áreas desmatadas.

Em geral, o processo de detecção de padrões de mudanças não é uma tarefa trivial, principalmente em grandes escalas, como é o caso da região Amazônica. Desta forma, é fundamental desenvolver ferramentas que permitam analisar padrões de mudança de maneira eficiente para tratar grandes volumes de dados. Um exemplo é o sistema GeoDMA (Geographic Data Mining Analyst) (KORTING, 2012), baseado em técnicas de mineração de dados e processamento de imagens para análises de padrões de uso e cobertura do solo (<http://www.dpi.inpe.br/geodma/>). A versão mais recente deste sistema é baseada principalmente em classificação por árvores de decisão. Neste sentido, existe um grande interesse em estendê-lo de forma a incluir novos classificadores e outras ferramentas de análise espaço-temporal para facilitar a identificação de padrões de mudança.

As contribuições deste trabalho incluem:

- (a) Uma comparação entre diferentes técnicas de classificação, desde as mais clássicas, e.g. Bayes Ingênuo e Vizinho Mais Próximo (THEODORIDIS; KOUTROUMBAS, 2008; THEODORIDIS et al., 2010), até as mais populares e recentes, e.g. Máquina de Vetores de Suporte (BOSER; GUYON; VAPNIK, 1992; CORTES; VAPNIK, 1995; CHANG; LIN, 2011) e Floresta de Caminhos Ótimos (PAPA, 2008; PAPA et al., 2012). Esta comparação foi baseada em uma validação cruzada *k*-fold (DUDA; HART; STORK, 2001) para proporcionar uma referência preliminar para estender, por exemplo, o sistema GeoDMA (KORTING, 2012) com um novo classificador.
- (b) Uma estratégia de classificação baseada em regiões (BLASCHKE; JOHANSEN; TIEDE, 2011) obtidas por segmentação, e.g. via Baatz (BAATZ; SCHAPE, 2000), utilizando perfis EVI para caracterizar diferentes temas do Projeto TerraClass, e.g. Agricultura. Idealmente, as regiões obtidas por este processo de segmentação correspondem a agrupamentos com perfis homogêneos para os índices de vegetação, reduzindo possíveis ruídos e o número de elementos a serem classificados. Atualmente, a classificação do TerraClass é realizada predominantemente por interpretação visual, cujo resultado depende fortemente da experiência de um especialista. Uma abordagem deste tipo pode proporcionar uma automatização de parte do processo de classificação.

O texto está organizado da seguinte maneira. A Seção 2 contextualiza o trabalho, exibindo uma breve descrição dos classificadores abordados e da estratégia baseada em regiões. A metodologia adotada para permitir uma comparação com diferentes classificadores é detalhada na Seção 3. Na Seção 4, os resultados experimentais são exibidos e interpretados. Finalmente, algumas conclusões são exibidas na Seção 5.

## 2. Trabalhos relacionados

Em reconhecimento de padrões, uma maneira de atribuir classes a objetos é através de uma classificação supervisionada. Neste tipo de classificação, as classes são conhecidas a priori. Por exemplo, a partir de dados rotulados de acordo com cada classe, é possível “treinar” um

classificador de maneira a construir um modelo de classificação.

Na literatura, existem diferentes classificadores seguindo esta estratégia. Dentre os mais simples, destacam-se o algoritmo de Bayes e o método do vizinho mais próximo. Referências sobre estas e outras técnicas comumente usadas podem ser encontradas em (THEODORIDIS; KOUTROUMBAS, 2008; THEODORIDIS et al., 2010).

No caso em que a distribuição dos dados e os parâmetros envolvidos são conhecidos, o método de Bayes é um classificador ótimo no sentido de minimizar o erro de classificação. Na prática, estas informações não são conhecidas e é comum assumir um formato Gaussiano para a distribuição, cujos parâmetros devem ser estimados. Além disso, especialmente para dados com alta dimensionalidade, costuma-se aplicar o método ingênuo, assumindo independência entre as variáveis (*Naive Bayes*, NB).

Para a estratégia baseada no vizinho mais próximo (*Nearest-Neighbor*, NN), a ideia é adotar uma métrica no espaço das características e classificar os dados de entrada de acordo com um protótipo, e.g. um padrão rotulado, associado à distância mínima.

Uma outra abordagem bastante popular é baseada em máquina de vetores de suporte (*Support Vector Machine*, SVM) (BOSER; GUYON; VAPNIK, 1992; CORTES; VAPNIK, 1995). Apesar de ser originalmente desenhada para uma classificação binária, é possível compor várias máquinas para tratar o problema com mais de duas classes. O SVM vem obtendo resultados satisfatórios para uma variedade de problemas e pode ser encontrado livremente na Internet. Por exemplo, o sítio do LIBSVM (CHANG; LIN, 2011) (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>) disponibiliza um conjunto de ferramentas baseadas no SVM, além de um guia prático para usuários iniciantes, onde é possível aplicar o método mesmo sem ter um domínio da teoria.

Alternativamente, é desejável aplicar métodos de classificação que naturalmente funcionam para duas ou mais classes, evitando assim a necessidade de uma composição de vários classificadores binários, diminuindo assim o tempo de execução. Por exemplo, o classificador baseado em caminhos ótimos (*Optimum Path Forest*, OPF) (PAPA, 2008; PAPA et al., 2012) (<http://www.ic.unicamp.br/~afalcao/libopf/>) funciona para duas ou mais classes. Este método é baseado em grafos. A ideia central é explorar uma conexidade no espaço de características, em que os padrões de entrada são classificados de acordo com um protótipo que oferece um caminho de menor custo. Este método é extremamente rápido e independe de parâmetros.

No contexto de classificadores, para o caso particular de reconhecimento de padrões em imagens, é possível adotar duas estratégias básicas: classificação baseada em pixels ou em regiões, e.g. resultantes de segmentação. De acordo com a literatura (BLASCHKE; JOHANSEN; TIEDE, 2011), em geral, abordagens baseadas em regiões são mais robustas, especialmente quanto à variação espectral devido à presença de ruídos e variação de iluminação. Além disso, agrupar pixels em regiões durante a fase de pré-processamento permite diminuir significativamente o número de protótipos e de padrões a serem classificados, reduzindo o tempo de processamento. O foco deste trabalho é sobre classificação baseada em regiões, envolvendo diferentes técnicas de classificação.

Para obter as regiões, em particular, o algoritmo de segmentação Baatz (BAATZ; SCHAPE, 2000) é popularmente aplicado, e.g. para possibilitar uma análise multirresolução de imagens. Como exemplos de sistemas utilizando este algoritmo, pode-se citar o eCognition (LANG; TIEDE, 2007), o InterIMAGE (COSTA et al., 2010) e o GeoDMA (KORTING, 2012). Neste tipo de análise, os pixels são agrupados em regiões homogêneas de acordo com as informações espectrais, e.g. cores. Neste trabalho, para comparar os diferentes classificadores, ao invés de usar bandas espectrais representando cores, as regiões foram obtidas através de perfis EVI

(*Enhanced Vegetation Index*). Neste caso, as regiões representam agrupamentos com perfis EVI homogêneos.

Os sistemas citados são predominantemente baseados em árvores de decisão e/ou em regras definidas pelo usuário. Dos três, apenas o eCognition, que é um sistema comercial, oferece também o classificador NN. O ENVI (ITT, 2008) é um exemplo de sistema comercial que oferece outros classificadores, e.g. *k*-Vizinhos Mais Próximos e SVM. Uma comparação entre diferentes classificadores pode proporcionar uma referência para possíveis extensões de sistemas abertos, como o GeoDMA e o InterIMAGE.

### 3. Metodologia

O alvo de estudo foi o estado do Mato Grosso. Nesta região existem extensas áreas com concentração de agricultura, pastagem e floresta (HUETE et al., 2002). Para identificar as classes correspondentes a diferentes tipos de uso de solo, foi utilizada uma sequência de imagens do sensor MODIS, adquiridas em instantes distintos. Os dados temporais correspondem a índices de vegetação EVI cobrindo todo o estado. Especificamente, foram utilizados perfis anuais como vetores de características. Cada perfil anual foi extraído da composição EVI de 8 dias do MODIS/Terra, resultando 45 valores/instantes para cada perfil ( $365 / 8 = 45,625$ ). A Figura 1 ilustra uma sequência de imagens, uma para cada instante. Para cada imagem, cada pixel corresponde a um valor EVI. Desta forma, empilhando a sequência de imagens resulta em perfis EVI, em que cada perfil inicia em um pixel da primeira imagem da sequência.

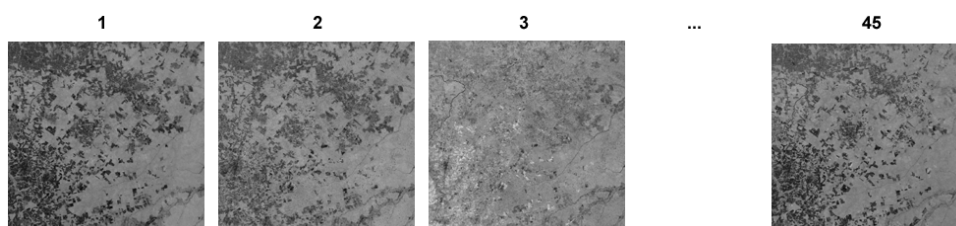


Figura 1: Sequência de imagens para compor perfis EVI: cada imagem corresponde a um instante no tempo, com pixels representando valores EVI naquele instante.

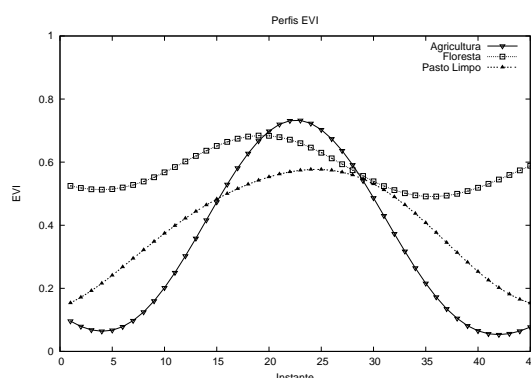


Figura 2: Exemplos de perfis EVI: floresta, agricultura e pasto limpo.

Para diminuir a quantidade de dados ruidosos, principalmente devido à presença de nuvens, os perfis EVI foram filtrados através do método de HANTS (*Harmonic Analysis of Time Series*) (ROERINK; MENENTI, 2000), resultando em curvas “suaves”, conforme ilustrado na Figura 2. Por exemplo, o perfil EVI anual da agricultura geralmente apresenta uma tendência

de crescimento durante a época de plantio e de desenvolvimento da cultura, seguida por uma tendência de queda correspondente à época de colheita. Já o perfil anual de uma floresta, em geral, não apresenta tendência, apenas uma oscilação aleatória em torno de um valor médio. O perfil de pasto limpo é um exemplo de “meio termo” entre o perfil da agricultura e da floresta. Em particular, este terceiro perfil ilustra o fato de que na prática, uma classificação perfeita é extremamente difícil de ser obtida devido à presença de imprecisões, principalmente nos dados rotulados, e principalmente às similaridades entre perfis pertencentes a classes distintas.

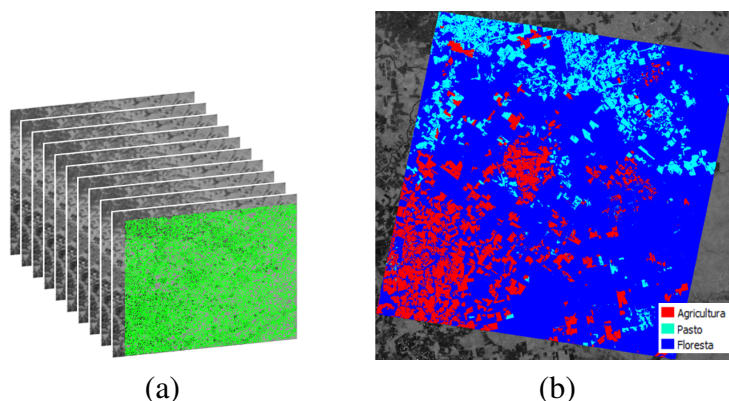


Figura 3: (a) Pilha de imagens representando uma composição EVI e sua correspondente segmentação (topo). (b) Recorte envolvendo a região órbita/ponto 226/68, com predomínio de agricultura, pasto e floresta, conforme mapa com os temas do TerraClass.

Dados rotulados foram obtidos através de mapas fornecidos pelo TerraClass para o ano de 2008 (imagens Landsat). Neste caso, houve uma necessidade de integrar dados obtidos por fontes diferentes: imagens Landsat (resolução espacial = 30 m) e MODIS (resolução espacial = 250 m). Desta forma, a classificação foi baseada em regiões (BLASCHKE; JOHANSEN; TIEDE, 2011) para agregar robustez ao processo de classificação com relação a ruídos e imprecisões nos dados. Baseando-se nos perfis anuais, recortes da composição (pilha de imagens) EVI foram segmentados pelo método Baatz (BAATZ; SCHAPE, 2000). Com o objetivo de obter regiões com perfis homogêneos, foram usados os seguintes parâmetros: escala = 100, compacidade = 0,5, cor = 0,9 e peso 1 para cada banda (instantes 1 a 45). A Figura 3(a) ilustra uma composição EVI e sua respectiva segmentação, exibida no topo da pilha. Por exemplo, aplicando a segmentação sobre um recorte com dimensões 885x839, envolvendo a região órbita/ponto 226/68, houve uma redução de  $885 \times 839 = 742515$  pontos para 12714 regiões. Para os experimentos, considerou-se o perfil médio de cada região.

Os diferentes temas do TerraClass incluem áreas de agricultura, pasto, regeneração, floresta secundária, hidrografia, área urbana, etc. Por exemplo, a região na órbita/ponto 226/68, exibida na Figura 3(b), tem um predomínio de agricultura, floresta e pasto limpo. Para simplificar a análise, os experimentos utilizaram duas classes: Agricultura e Não-agricultura. A Agricultura como tema de interesse, *versus* o restante, e.g. Pasto, Floresta, etc.

Os dados rotulados foram obtidos da seguinte maneira. Através dos mapas fornecidos pelo TerraClass para o ano de 2008, e dos perfis EVI correspondentes ao período entre início de junho de 2007 e final de junho de 2008, cada região foi atribuída com o rótulo correspondente à classe de maior ocorrência. Para a região na órbita/ponto 226/68, foram rotuladas 8191 regiões (das 12714 obtidas pela segmentação): 1712 regiões rotuladas como Agricultura e 6479 regiões rotuladas como Não-agricultura. Neste caso, há um desbalanceamento nos dados rotulados. Para balanceá-los, o excesso de protótipos de Não-agricultura foi descartado de maneira aleatória, e.g. após o balanceamento a região órbita/ponto 226/68 ficou com 1712

Tabela 1: Valores médios de Acurácia, Sensibilidade e Especificidade, com os respectivos desvios padrão, obtidos a partir da validação cruzada  $k$ -fold para a região órbita/ponto 226/68.

Método	226/68		
	Acurácia	Sensibilidade	Especificidade
NB	0,8473 +- 0,0194	0,8581 +- 0,0324	0,8365 +- 0,0243
NN	0,8347 +- 0,0207	0,8253 +- 0,0266	0,8441 +- 0,0230
DT	0,8201 +- 0,0192	0,8096 +- 0,0386	0,8306 +- 0,0162
OPF	0,8280 +- 0,0200	0,8107 +- 0,0268	0,8452 +- 0,0223
<b>SVM</b>	<b>0,8823 +- 0,0176</b>	<b>0,9024 +- 0,0235</b>	<b>0,8622 +- 0,0254</b>

regiões para cada classe, totalizando 3424 regiões rotuladas.

Seguindo uma estratégia de classificação supervisionada, um conjunto de dados rotulados pode ser dividido em dois subconjuntos: um conjunto de treinamento e outro de teste. A ideia é criar um modelo baseado nos dados de treinamento para classificar dados de entrada. Através dos dados de teste, é possível avaliar o desempenho do modelo. Para os experimentos, utilizou-se a validação cruzada  $k$ -fold (DUDA; HART; STORK, 2001). Neste tipo de validação, o conjunto rotulado é dividido em  $k$  subconjuntos. Para cada subconjunto, a ideia é usá-lo como teste e o restante como treinamento ( $k - 1$  subconjuntos), totalizando  $k$  avaliações.

A comparação experimental envolveu os seguintes métodos: SVM, OPF, NB, NN e DT. Para testar o método SVM, utilizou-se a biblioteca LIBSVM (CHANG; LIN, 2011). Para o método OPF, utilizou-se a biblioteca disponível pelos autores (PAPA, 2008). Ambas disponibilizam ferramentas para particionar um conjunto rotulado de dados (e.g. em treinamento e teste), treinar o classificador, e classificar dados de entrada (e.g. usando os dados de teste para avaliar o desempenho do modelo construído durante a fase de treinamento). Além disso, o SVM disponibiliza uma ferramenta para estimar os parâmetros do modelo, enquanto que o OPF é livre de parâmetros. Já os métodos NB e NN foram implementados em MatLab, seguindo (THEODORIDIS; KOUTROUMBAS, 2008; THEODORIDIS et al., 2010). Para o NB, assumiu-se uma distribuição Gaussiana para os dados, e estimação por máxima verossimilhança para os parâmetros (média e desvio padrão). Para o NN, assumiu-se a distância Euclidiana entre os perfis anuais EVI. Finalmente, para proporcionar uma referência para sistemas baseados em classificação com árvores de decisão (DT), foram incluídos resultados da função `classregtree` do Matlab.

#### 4. Resultados Experimentais

As Tabelas 1, 2 e 3 exibem os índices médios de Acurácia, de Sensibilidade e de Especificidade, com os correspondentes desvios, obtidos pelos diferentes métodos para diferentes regiões (órbita/ponto). A Acurácia é definida como o número de regiões classificadas corretamente dividido pelo total de regiões classificadas. A Sensibilidade é definida pela acurácia restrita à classe Agricultura. A Especificidade é definida pela acurácia restrita à classe Não-agricultura. As médias e os desvios foram obtidos seguindo uma validação cruzada  $k$ -fold, para  $k = 10$ . Desta forma, as médias foram obtidas de  $k$  avaliações. Em particular, a região órbita/ponto 226/69 apresentou um predomínio de Agricultura, Floresta e Não-floresta nos temas do TerraClass. Neste caso, um número significativo de regiões associadas a Não-floresta apresentou perfil similar ao de Agricultura, reduzindo os índices de Especificidade. A Figura 4 é dedicada à classificação de Agricultura, exibindo os gráficos com os índices de Sensibilidade obtidos para cada uma das  $k$  avaliações. Os resultados indicam que o SVM obteve a melhor performance dentre os métodos testados.

Tabela 2: Valores médios de Acurácia, Sensibilidade e Especificidade, com os respectivos desvios padrão, obtidos a partir da validação cruzada  $k$ -fold para a região órbita/ponto 227/68.

Método	227/68		
	Acurácia	Sensibilidade	Especificidade
NB	0,8496 +- 0,0223	0,8705 +- 0,0211	0,8286 +- 0,0318
NN	0,8254 +- 0,0256	0,8169 +- 0,0294	0,8339 +- 0,0337
DT	0,8259 +- 0,0276	0,8243 +- 0,0303	0,8274 +- 0,0398
OPF	0,8164 +- 0,0240	0,8015 +- 0,0276	0,8312 +- 0,0342
<b>SVM</b>	<b>0,8896 +- 0,0120</b>	<b>0,9166 +- 0,0176</b>	<b>0,8625 +- 0,0161</b>

Tabela 3: Valores médios de Acurácia, Sensibilidade e Especificidade, com os respectivos desvios padrão, obtidos a partir da validação cruzada  $k$ -fold para a região órbita/ponto 226/69.

Método	226/69		
	Acurácia	Sensibilidade	Especificidade
NB	0,7716 +- 0,0134	0,8097 +- 0,0182	0,7335 +- 0,0298
NN	0,7371 +- 0,0197	0,7270 +- 0,0349	<b>0,7471 +- 0,0218</b>
DT	0,7300 +- 0,0198	0,7236 +- 0,0242	0,7364 +- 0,0359
OPF	0,7275 +- 0,0219	0,7120 +- 0,0345	0,7429 +- 0,0247
<b>SVM</b>	<b>0,8134 +- 0,0149</b>	<b>0,9104 +- 0,0170</b>	0,7163 +- 0,0284

## 5. Conclusão

O SVM é originalmente desenhado para problemas com duas classes, mas é possível combinar vários classificadores SVMs para tratar mais de duas classes. Durante os experimentos, o fator predominante no tempo de processamento do SVM foi a fase de estimação de parâmetros. Já o método OPF é naturalmente multiclasse e independe de parâmetros, sendo bastante rápido na prática. Mas, apesar das vantagens oferecidas pelo OPF, os resultados experimentais indicaram que o SVM foi superior em relação ao OPF e demais métodos. Os resultados concordam com a literatura, em que o SVM vem obtendo ótimos resultados de classificação em diferentes aplicações, sendo um potencial candidato para ser incluído em sistemas abertos como o GeoDMA e o InterIMAGE.

**Agradecimentos:** À FAPESP pelo apoio financeiro (proc. 2010/18854-8). Ao prof. Bernardo Rudorff pela disponibilização dos dados EVI usados nos experimentos. Ao prof. João Papa pela ajuda com o OPF e pelas sugestões.

## Referências

- BAATZ, M.; SCHAPE, A. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In: ANGEWANDTE GEOGRAPHISCHE INFORMATIONSVERRARBEITUNG XII. [S.l.], 2000.
- BLASCHKE, T.; JOHANSEN, K.; TIEDE, D. Object-based image analysis for vegetation mapping and monitoring. In: ADVANCES IN ENVIRONMENTAL REMOTE SENSING: SENSORS, ALGORITHMS, AND APPLICATIONS. [S.l.]: CRC Press, Taylor and Francis, 2011.
- BOSER, B. E.; GUYON, I.; VAPNIK, V. A training algorithm for optimal margin classifiers. In: PROCEEDINGS OF THE FIFTH ANNUAL WORKSHOP ON COMPUTATIONAL LEARNING THEORY. [S.l.]: ACM Press, 1992. p. 144–152.
- CHANG, C.-C.; LIN, C.-J. Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, v. 2, n. 3, p. 1–27, 2011.
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine Learning*, v. 20, n. 3, p. 273–297, 1995.

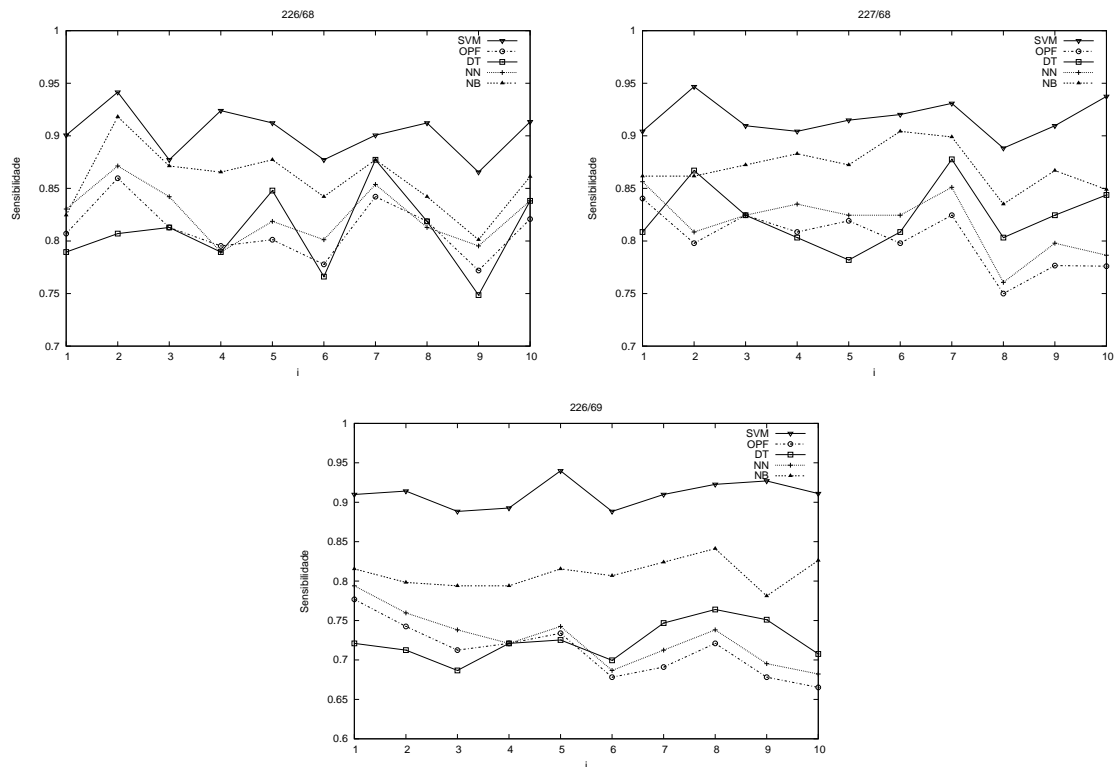


Figura 4: Agricultura: Gráficos com os índices de sensibilidade individuais para cada avaliação  $i = 1 \dots k$  da validação cruzada  $k$ -fold.

COSTA, G. et al. Knowledge-based interpretation of remote sensing data with the interimage system: major characteristics and recent developments. In: GEOBIA. [S.l.], 2010.

DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern Classification*. Second. [S.l.]: Wiley, 2001.

HUETE, A. et al. Overview of the radiometric and biophysical performance of the modis vegetation indices. *Remote Sensing of Environment*, v. 83, n. 1–2, p. 195–213, 2002.

ITT. Envi feature extraction module user's guide. In: GILCHING, GERMANY: EXELIS VISUAL INFORMATION SOLUTIONS. [S.l.], 2008.

KORTING, T. S. *GeoDMA: a toolbox for data mining, object-based and multi-temporal analysis of remote sensing imagery*. Orientadores: L. M. G. Fonseca, G. Câmara. Tese (Doutorado) — INPE, 2012.

LANG, S.; TIEDE, D. Definiens developer. *GIS Business* 9, p. 34–37, 2007.

PAPA, J. P. *Classificação supervisionada de padrões utilizando floresta de caminhos ótimos*. Orientador: A. X. Falcão. Tese (Doutorado) — Unicamp, 2008.

PAPA, J. P. et al. Efficient supervised optimum-path forest classification for large datasets. *Pattern Recognition*, v. 45, n. 1, p. 512–520, 2012.

ROERINK, G. J.; MENENTI, M. Reconstructing cloudfree ndvi composites using fourier analysis of time series. *International Journal of Remote Sensing*, v. 21, n. 9, p. 1911–1917, 2000.

THEODORIDIS, S.; KOUTROUMBAS, K. *Pattern Recognition*. Fourth. [S.l.]: Academic Press, 2008.

THEODORIDIS, S. et al. *Introduction to Pattern Recognition: A Matlab Approach*. First. [S.l.]: Academic Press, 2010.