Classificação de pastagens cultivadas e formações campestres nativas no Cerrado brasileiro a partir da análise de séries temporais extraídas de índices EVI do sensor MODIS

Wanderson Santos Costa[1]
Leila Maria Garcia Fonseca[1]
Thales Sehn Körting[1]

[1] Instituto Nacional de Pesquisas Espaciais – INPE
Caixa Postal 515 – 12245-970 – São José dos Campos - SP, Brasil
wscosta, leila, thales@dpi.inpe.br

Abstract. One of the most biodiverse regions on the planet, the Brazilian Cerrado has an area of approximately 2 million km^2 and it is the second largest biome in Brazil. Among the land cover modifications in the biome, over one fourth of its area has been changed into cultivated pastures in the last few years. Categorizing types of land use and cover in the Cerrado and its native formations is important for protection policy and monitoring of the Brazilian Cerrado. Based on remote sensing techniques, this work aims at developing a methodology to map pasture and native grassland areas in the biome. Data related to EVI vegetation indices obtained by MODIS images were used to perform image classification. This study encompasses a Cerrado area that comprises a region of Serra da Canastra National Park and neighboring regions, that contains all targets of interest. Support Vector Machines, Decision Trees and Random Forests algorithms were compared, and the results showed that the analysis of different attributes extracted from EVI indices can aid in the classification process. As a means to distinguish grassland and cultivated pasture zones, we obtained accuracies up to 85,96% in the study area,identifying data and attributes required to identify these areas by remote sensing images.

Keywords: Cerrado, image processing, data mining .

1. Introdução

Mais da metade da área do Cerrado brasileiro tem sido transformada, principalmente, em áreas de pastagem e agricultura, perdendo cerca de 1 milhão de km^2 de sua vegetação original (MACHADO et al., 2004). As áreas de agricultura cobrem mais de 100.000 km^2 e as de pastagem cultivada superam 500.000 km^2, enquanto que as áreas preservadas envolvem apenas cerca de 33.000 km^2. Esta situação representa um alto custo ambiental, ocasionando perda de biodiversidade, erosão do solo, degradação das formações vegetais do bioma, poluição da água, mudanças nos eventos de fogo típicos do bioma, instabilidade do ciclo do carbono e prováveis modificações climáticas regionais (KLINK; MACHADO, 2005).

Segundo Walter (2006), dentre as formações vegetais existentes no bioma, encontram-se as formações campestres, que se referem a regiões com predominância de espécies herbáceas e algumas arbustivas, sem ou com pouca ocorrência de árvores na paisagem (RIBEIRO; WALTER, 2008). Considerando que estas fitofisionomias são formações naturais, elas são passíveis de recuperação caso sejam degradadas. Neste caso, o monitoramento deste tipo de vegetação, por meio de imagens de satélite, pode auxiliar na recuperação da sua integridade física, química e biológica assim como na capacidade produtiva da região degradada. Consequentemente, mapas de uso e cobertura do Cerrado, obtidos a partir imagens de satélites e técnicas de sensoriamento remoto, podem ser usados para monitorar estas áreas.

Existem alguns estudos, como o de Ferreira et al. (2013b), que usam técnicas de análise multitemporal baseada em índices de vegetação para classificar áreas de pastagem cultivada no Cerrado. Entretanto, as áreas transformadas em pastagem, de forma similar às formações

campestres supracitadas, podem variar de formações com predominância de gramíneas a locais que apresentam dominância de espécies arbustivas, por exemplo (EMBRAPA; INPE, 2011). Desta forma, o mapeamento de áreas de pastagem cultivada e de formações campestres nativas no Cerrado, por meio de imagens de satélites, pode se tornar uma tarefa difícil quando usada apenas a informação espectral dos alvos (SANO et al., 2008). Dentro deste contexto, este trabalho tem como objetivo classificar e criar mapas temáticos de áreas de formações campestres nativas e regiões de pastagens cultivadas no Cerrado brasileiro por meio de técnicas de geoprocessamento e da implementação de algoritmos, utilizando métodos de análise de padrões e mineração de dados em dados de séries temporais de imagens. Algoritmos de árvores de decisão (DT), florestas aleatórias (RF) e *Suppport Vector Machine* (SVM) foram comparados na análise dos resultados.

O trabalho foi organizado da seguinte forma: na Seção 2 é apresentada uma breve descrição do Cerrado brasileiro, com ênfase nas formações vegetais nativas e áreas antrópicas de interesse. Os dados e a metodologia descritos no trabalho estão detalhados na Seção 3 . Na Seção 4 são apresentados os experimentos realizados e uma discussão sobre os progressos obtidos neste estudo. Finalmente, na Seção 5 são apresentadas as considerações finais e trabalhos futuros.

2. Cerrado: Campos Nativos e Pastagens Cultivadas

O termo Cerrado é comumente empregado para indicar o conjunto de ecossistemas (savanas, campos e matas) que ocupa a região central do Brasil (EITEN, 1977) e pode ser entendido como uma formação composta por campos graminosos coexistentes com árvores e arbustos esparsos. Correspondente a uma das savanas mais úmidas do planeta, o bioma apresenta um clima sazonal, com estações seca e chuvosa bem definidas. Várias propostas de classificação das formações vegetais do bioma são apresentadas na literatura (WALTER, 2006). Ribeiro e Walter (2008), por exemplo, descrevem o Cerrado como uma região composta de formações florestais, savânicas e campestres. Em particular, as formações campestres referem-se a regiões com predominância de espécies herbáceas e algumas arbustivas, sem a ocorrência de árvores na paisagem. Dentre as formações vegetais do bioma, este trabalho tem como foco as formações campestres nativas no Cerrado, subdivididas em Campo Limpo, Campo Sujo e Campo Rupestre.

O Campo Limpo é uma fitofisionomia com predominância herbácea e contém poucos arbustos, com a inexistência de árvores na região. É localizado com frequência nas chapadas, olhos d'água e encostas, com locais que podem apresentar diversas posições topográficas e variações no grau de profundidade e fertilidade do solo. Áreas de Campo Sujo contêm, em seu estrato arbustivo-herbáceo, espécies de arbustos e subarbustos muitas vezes compostas por plantas menos desenvolvidas das espécies arbóreas do Cerrado Sentido Restrito. Podem ser encontradas em solos rasos, com presença eventual de pequenos afloramentos de rocha de pouca extensão ou mesmo em terrenos profundos e de pouca fertilidade. A fitofisionomia de Campo Rupestre agrupa paisagens em microrrelevos com espécies típicas, geralmente ocupando trechos de afloramento rochosos em altitudes superiores a 900 metros, ocasionalmente a partir de 700 metros, com a presença eventual de arvoretas raquíticas. São, em geral, regiões com solos ácidos e pobres em nutrientes, onde existem altas variações de temperatura e ventos constantes (RIBEIRO; WALTER, 2008).

Por outro lado, em relação às áreas antropizadas do Cerrado, estudos mostram que as mudanças do uso do solo no Cerrado ocorrem com maior intensidade do que na região Amazônica (SANO; BARCELLOS; BEZERRA, 2000) e mais da metade da vegetação original do Cerrado foi transformada em áreas de pastagem, agricultura e outros usos (MACHADO et al., 2004). Das regiões transformadas no Cerrado, as áreas de Pastagem Cultivada correspondem a cerca de 500.000 km^2 (MACHADO et al., 2004), ocupando cerca de um quarto da sua extensão.

De acordo com Ferreira et al. (2013b), no mínimo, 50 % das áreas de pastagem plantada no Cerrado já se encontram severamente degradadas, ocasionando o aumento da erosão e a perda da fertilidade do solo, a redução da biomassa e a predominância de espécies estrangeiras, principalmente as espécies de gramíneas africanas do gênero *Brachiaria*. Com isso, a recuperação destas áreas degradadas pode auxiliar no aumento da renda para os produtores e pode-se reduzir o impacto ambiental no Cerrado por meio da diminuição da erosão, da emissão de dióxido de carbono e da abertura de novas áreas para pastagem (CHAVES et al., 2001).

A identificação das regiões de pastagem é difícil porque a degradação dos pastos pode, por exemplo, influenciar na porcentagem da cobertura vegetal e na resposta dos índices de vegetação. Confusão e incertezas podem ocorrer quando as pastagens são manejadas de forma inadequada, uma vez que podem aparecer espécies invasoras ou mesmo o renascimento de espécies de arbustos e árvores nativas nestas regiões (FERREIRA et al., 2013b).

Em decorrência das grandes extensões de áreas, da periodicidade marcante da vegetação natural e da confusão espectral entre classes, a necessidade de identificar e discriminar áreas de pastagens cultivadas e pastagens naturais (formações campestres) é um grande desafio (SANO et al., 2008). Consequentemente, para melhorar a discriminação destes alvos é necessário usar dados temporais e de campo, explorando a vantagens de o bioma apresentar estações bem definidas (FERREIRA et al., 2013a).

3. Materiais e Métodos

3.1. Área de estudo

A área de estudo faz parte do complexo do Cerrado, e apresenta regiões naturais e antrópicas, com áreas de pastagem cultivada. Nesta região, ilustrada na Figura 1, está presente o Parque Nacional da Serra da Canastra e regiões vizinhas, onde são encontradas todas as fitofisionomias do grupo das formações campestres nativas. Esta região foi escolhida por apresentar regiões de Campo Nativo e Pastagem Cultivada.

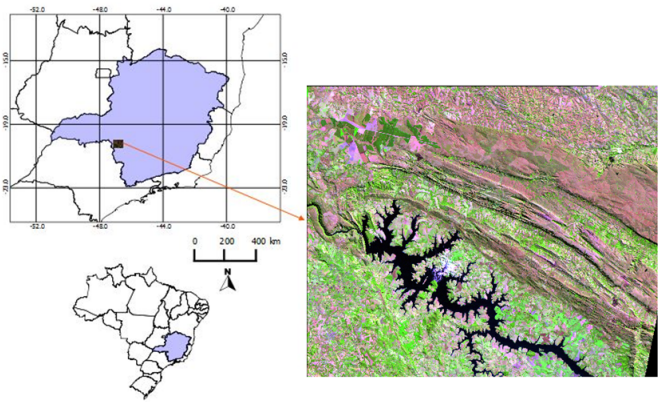

Figura 1: Área de estudo do trabalho, localizada no sudoeste do estado de Minas Gerais. Recorte de uma imagem Landsat TM-5 (R5B4G3) da região.

3.2. Dados e Técnicas

Os dados e as técnicas que foram utilizadas para a geração de mapas temáticos como plano de informação resultante da classificação proposta neste trabalho estão esquematizados no modelo OMG-T exposto a seguir (Figura 2). O modelo apresenta uma descrição dos tipos de dados de entrada e as etapas de processamento dos dados que foram usados para o

desenvolvimento desta pesquisa.

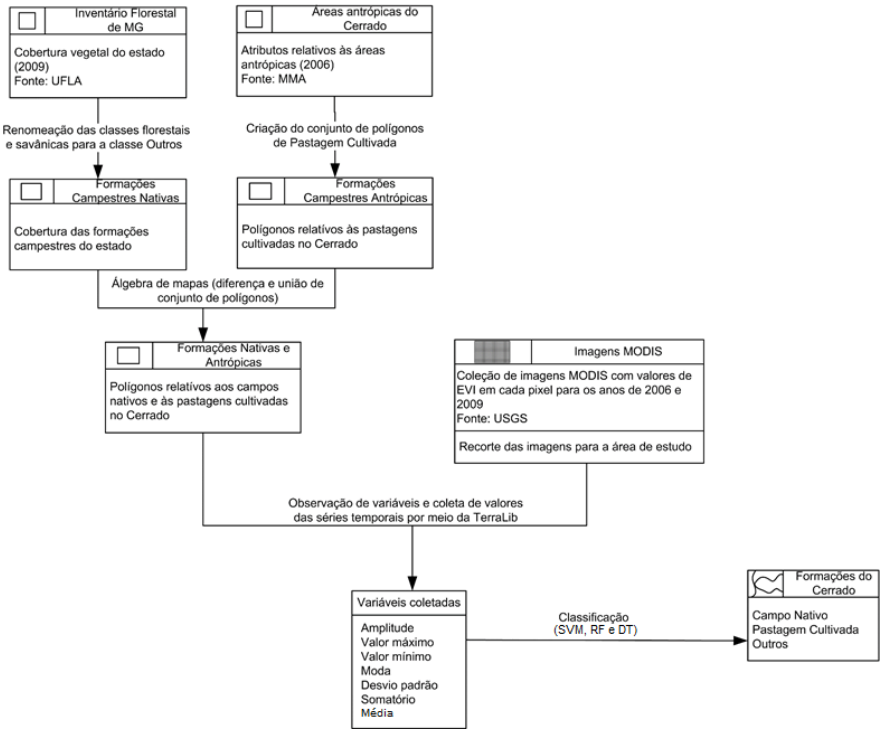

Figura 2: OMT-G do trabalho proposto

O mapa de cobertura vegetal do Cerrado fornecido pelo Inventário Florestal de Minas Gerais do ano de 2009 foi utilizado para a obtenção das formações campestres nativas. O *shapefile* contém polígonos para todo o estado de Minas Gerais e contém duas classes que representam as formações campestres nativas do Cerrado: a classe Campo, que combina as fitofisionomias de Campo Limpo e Campo Sujo; e a classe Campo Rupestre, definida pela fitofisionomia homônima. Em relação às pastagens cultivadas do Cerrado, o Ministério do Meio Ambiente (MMA) oferece um *shapefile*, relativo ao ano de 2006, que distingue a cobertura antrópica da natural no bioma e faz a discriminação das áreas antrópicas do Cerrado, contendo polígonos rotulados como áreas de pastagem cultivada.

Ambos os mapas de referência resultantes foram gerados utilizando as ferramentas de álgebra de mapas disponíveis no software TerraView (INPE, 2014) juntamente o plugin GeoDMA (*Geographic Data Mining Analyst*), desenvolvido por Körting, Fonseca e Câmara (2013), conforme descrito na Figura 2. Com isso, criaram-se dois arquivos *shapefile* sem polígonos sobrepostos: um com polígonos associados a duas classes: Campo Nativo e Outros para o ano de 2009; e outro com polígonos categorizados como Pastagem Cultivada para o ano de 2006. Estes polígonos são usados nas etapas de extração de atributos e treinamento, e também como referência para avaliar os resultados da classificação.

A presença de duas estações bem definidas no Cerrado pode auxiliar na detecção de mudanças nas coberturas vegetais a partir da observação de séries temporais. Para analisar esta situação, foram utilizadas as imagens do sensor MODIS, com correção atmosférica e resolução espacial de 250 m, para os períodos seco e úmido da região, utilizando dados temporais dos índices de vegetação EVI. Especificamente, foram utilizados perfis anuais como vetores de características para 2006 e 2009. Cada perfil anual foi extraído da composição EVI de 16 dias do MODIS, resultando em 23 instantes para cada perfil. Cada pixel das imagens corresponde a um valor de índice de vegetação (EVI), de modo que os valores ordenados desta sequência de

imagens resulta em uma série temporal para cada pixel.

Para a extração dos atributos preditivos, cada amostra foi considerada como o conjunto de informações relativas a um polígono do conjunto de regiões dos *shapefiles* criados a partir dos mapas de referência. Em relação às imagens MODIS, para cada uma das 23 imagens dos índices EVI, os atributos relacionados com os índices de vegetação foram extraídos a partir da criação de algoritmos usando a biblioteca TerraLib 5. Foi necessário implementar um algoritmo para associar a média dos valores de todos os pixels (no caso, os índices de vegetação) para cada um dos polígonos. Os atributos dos índices de vegetação correspondem, assim, à média dos valores dos pixels dentro de cada polígono, para cada imagem EVI no instante $T_i, i = 1, 2, \cdots, 23$.

Para evitar o problema de pixels de borda, foi utilizada uma abordagem para definir os atributos de índices de vegetação para cada polígono. Usando a biblioteca TerraLib 5, foi implementado um algoritmo que define o valor do atributo EVI para cada polígono em cada uma das imagens T_1 a T_{23}, equivalente à média dos "pixels puros" do polígono em questão. A Figura 3 exibe dois exemplos de pixels no processo de decisão do algoritmo.

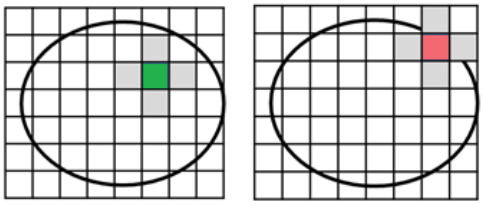

Figura 3: Para um polígono, o pixel em destaque da figura à esquerda é considerado um pixel puro, enquanto o da segunda figura não obedece à relação espacial definida no algoritmo e, portanto, não é levado em consideração.

Foi considerado, no cálculo do valor EVI para um polígono em cada uma das 23 imagens, apenas os pixels cujos seus vizinhos também pertencem à mesma classe, evitando, assim, que pixels de borda influenciem no cálculo da média dos valores de EVI de um dado polígono. A partir desta estratégia, consequentemente, polígonos muito pequenos, que estão inclusos em único pixel são desconsiderados do conjunto de amostras.

Com a geração da série temporal para cada um dos polígonos, o algoritmo elaborado neste trabalho também faz o cálculo e armazenamento das variáveis derivadas da série temporal extraída, sendo estas variáveis também utilizadas como atributos nos experimentos realizados. As variáveis extraídas a partir da série temporal são os valores de: mínimo, máximo, amplitude, soma, moda, média e desvio padrão.

Foi utilizada a estratégia de divisão percentual do conjunto de total de amostras, sendo 2/3 das amostras para o conjunto de treinamento e 1/3 para o conjunto de teste. Para o método SVM, utilizou-se a biblioteca LibSVM (CHANG; LIN, 2011), enquanto o *framework* Weka foi utilizado para os métodos de árvores de decisão (DT) e florestas aleatórias (RF) (WITTEN; FRANK, 2000). No processo de avaliação, as matrizes de confusão e os índices Kappa (COHEN, 1960) foram usados. A partir da matriz de confusão, podem-se calcular medidas estatísticas descritivas, como a acurácia dos resultados, para que a confusão entre classes possa ser claramente apresentada e entendida. No caso da classificação SVM, os cálculos da matriz de confusão e do índice Kappa global foram implementados a partir de algoritmos em linguagem R e C++.

É importante salientar que a presença inerente de ruídos nas imagens, tais como falhas nos dados temporais e cobertura por nuvens, podem restringir a utilização dos indicadores de fenologia vegetal para a classificação da cobertura do solo, podendo contribuir na redução da acurácia dos resultados (LAMBIN; LINDERMAN, 2006). Após utilizar e avaliar os resultados dos

classificadores, são construídos como saída do processo de classificação planos de informação referentes aos mapas temáticos para cada um dos experimentos classificados, sendo possível gerar uma camada de informação a partir da integração e processamento dos dados de entrada.

4. Resultados e Discussões

Para a geração e análise dos resultados, três experimentos foram realizados. Cada experimento usa um conjunto de atributos específico no processo de classificação, como mostra a Tabela 1.

Tabela 1: Resumo dos conjuntos de dados utilizados em cada experimento.

Experimento	Atributos preditivos utilizados
1	23 atributos (Série temporal)
2	Atributos extraídos a partir da série temporal
3	Série temporal e atributos extraídos a partir da série temporal

As classes de interesse utilizadas neste estudo foram: Campo Nativo, Pastagem Cultivada e Outros. Os métodos de classificação SVM, florestas aleatórias (RF) e árvores de decisão (DT) foram utilizados nos três experimentos. Além da matriz de confusão e acurácia, o índice Kappa também foi calculado para todos os resultados de classificação com o intuito de mensurar a concordância entre os resultados dos classificadores e o conjunto de referência.

A partir dos algoritmos implementados em TerraLib 5 utilizando a abordagem de pixels puros descrita na seção anterior, foi extraído um conjunto com um total de 508 amostras para realizar os três experimentos, onde cada amostra é relativa a atributos de um polígono. Foram separadas 337 amostras para o conjunto de treinamento e 171 amostras para o de teste. A Tabela 2 exibe os resultados obtidos a partir dos três classificadores.

Tabela 2: Resultados das classificações dos experimentos realizados

Experimento	Acurácia (%)			Índice Kappa		
	SVM	RF	DT	SVM	RF	DT
1	84,21	80,70	71,35	0,74	0,68	0,52
2	76,60	71,34	69,01	0,61	0,52	0,47
3	**85,96**	81,87	76,02	**0,77**	0,70	0,60

Em todos os três experimentos, o método SVM apresentou os melhores resultados da classificação, seguido do método de florestas aleatórias e o de árvores de decisão, respectivamente. No Experimento 1, avaliou-se o potencial das séries temporais das imagens EVI como único conjunto de atributos na classificação, a fim de analisar a contribuição das séries temporais em si. A melhor classificação (SVM) resultou em uma acurácia de 84,21% e índice Kappa de 0,74.

No Experimento 2, foi avaliada a contribuição das variáveis extraídas a partir de cada uma das séries temporais. Os atributos preditivos utilizados foram os valores de mínimo, máximo, amplitude, soma, moda, média e desvio padrão, todos também extraídos e implementados em um algoritmo em TerraLib 5. Os resultados utilizando apenas estas informações não implicaram em uma boa classificação, se comparada com o Experimento 1. O melhor valor de acurácia foi de apenas 76,60%, com um índice Kappa de 0,61.

Por fim, no Experimento 3 foi analisado se a combinação dos atributos do Experimento 1 e do Experimento 2 alcançaria melhores resultados. A partir da avaliação dos resultados,

pode-se perceber um pequeno aumento na acurácia em relação aos outros testes aplicados. O Experimento 3 utilizando SVM alcançou o melhor resultado das classificações realizadas e atingiu acurácia de 85,96% e índice Kappa de 0,77, apenas com um aumento na acurácia de 1,75% em relação ao Experimento 1, sem trazer um acréscimo mais significativo. O mapa temático gerado a partir da classificação para o Experimento 3 pode ser observado na Figura 4 para comparação com o conjunto verdade dos polígonos utilizados.

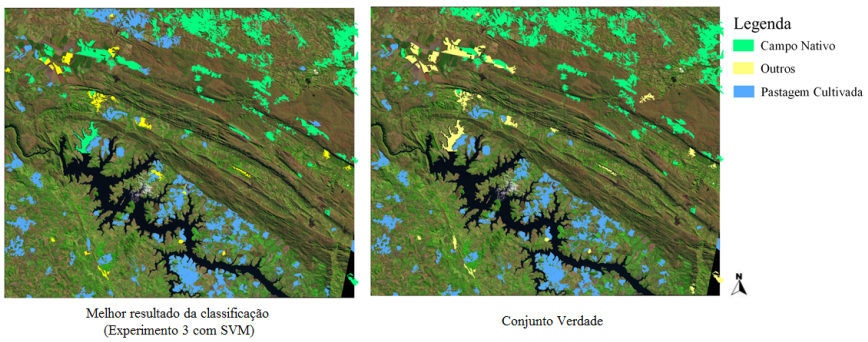

Figura 4: Mapas temáticos gerados a partir da melhor classificação e do conjunto verdade das amostras. Os planos de informação estão sobrepostos em uma imagem Landsat TM-5 (R5B4G3).

Do ponto de vista operacional, a criação de um modelo de classificação por SVM com os atributos do Experimento 3 apresentaram resultados similares ao Experimento 1, já que a melhora na acurácia foi pequena para o mapeamento das três classes. Além disso, pode-se considerar que, a partir da construção de mapas temáticos como saída do processo de classificação, o trabalho pode contribuir na integração de dados de sensoriamento remoto em sistemas de informação de geográfica, uma vez que foi possível gerar uma camada de informação a partir dos dados de entrada.

5. Conclusões e Trabalhos Futuros

O trabalho teve o intuito de classificar e criar mapas temáticos de áreas de formações campestres nativas e regiões de pastagens cultivadas no Cerrado brasileiro a partir da implementação de algoritmos e de técnicas de mineração de dados e geoprocessamento. Para isso, uma região do Parque da Serra da Canastra, situada em Minas Gerais, foi utilizada como área de estudo, uma vez que podem ser encontradas áreas correspondentes às classes de interesse nestas localizações.

O processo de análise de dados de sensoriamento remoto foi realizado com auxílio dos ambientes computacionais GeoDMA e Weka, além da biblioteca LibSVM. Algoritmos foram implementados na biblioteca TerraLib 5, em linguagem R e C++ para a classificação das amostras, avaliação e validação dos resultados, além da possibilidade de considerar apenas os pixels puros para o cálculo das amostras. Além da implementação dos algoritmos, várias operações de geoprocessamento foram utilizadas para que os planos de informações temáticos fossem gerados no final do processo de classificação, tais como álgebras de mapas, reclassificação de geo-objetos, recorte e registro de imagens.

Três experimentos foram conduzidos, combinando atributos preditivos relativos e derivados das séries temporais de índices EVI para os anos de 2006 e 2009 da área de estudo, sendo utilizados para a extração dos atributos preditivos, o treinamento e a avaliação do conjunto de amostras. O algoritmo SVM apresentou os melhores resultados na separação entre campo nativo e pastagens cultivadas para as amostras em questão. O melhor resultado obtido teve acurácia

de 85,96% e índice Kappa de 0,77, combinando todos os atributos referentes à série temporal e as variáveis que foram extraídas a partir desta. Como trabalhos futuros, pretende-se analisar o potencial dos atributos extraídos neste trabalho juntamente com outras fontes de dados de sensoriamento remoto de diferentes resoluções espaço-temporais.

Referências

CHANG, C.-C.; LIN, C.-J. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, ACM, v. 2, n. 3, p. 27, 2011.

CHAVES, J. M. et al. Uso da técnica de segmentação na identificação dos principais tipos de pastagens cultivadas do cerrado. In: *Anais*. São José dos Campos: INPE, 2001. p. 31–33.

COHEN, J. A. Coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, n. 20, p. 37–46, 1960.

EITEN, G. Delimitação do conceito de cerrado. *Arquivos do Jardim Botânico*, 1977.

EMBRAPA; INPE. *Levantamento de informações de uso e cobertura da terra na Amazônia*. Brasília, 2011.

FERREIRA, L. G. et al. Biophysical properties of cultivated pastures in the brazilian savanna biome: An analysis in the spatial-temporal domains based on ground and satellite data. *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 5, n. 1, p. 307–326, 2013.

FERREIRA, L. G. et al. Biophysical characteristics and fire occurrence of cultivated pastures in the brazilian savanna observed by moderate resolution satellite data. *International Journal of Remote Sensing*, Taylor & Francis, v. 34, n. 1, p. 154–167, 2013.

INPE. *TerraView*. São José dos Campos, SP: Divisão de Processamento de Imagens, 2014.

KLINK, C. A.; MACHADO, R. B. Conservation of the brazilian cerrado. *Conservation biology*, Wiley Online Library, v. 19, n. 3, p. 707–713, 2005.

KÖRTING, T. S.; FONSECA, L. M. G.; CÂMARA, G. Geodma–geographic data mining analyst. *Computers & Geosciences*, Elsevier, v. 57, p. 133–145, 2013.

LAMBIN, E. F.; LINDERMAN, M. Time series of remote sensing data for land change science. *Geoscience and Remote Sensing, IEEE Transactions on*, IEEE, v. 44, n. 7, p. 1926–1928, 2006.

MACHADO, R. B. et al. Estimativas de perda da área do cerrado brasileiro. *Conservation International do Brasil, Brasília*, 2004.

RIBEIRO, J. F.; WALTER, B. M. T. As principais fitofisionomias do bioma cerrado. In: *Cerrado: Ecologia e Flora*. Brasília: EMBRAPA, 2008. v. 1, p. 152–212.

SANO, E. E.; BARCELLOS, A. O.; BEZERRA, H. S. Assessing the spatial distribution of cultivated pastures in the brazilian savanna. *Pasturas Tropicales*, Centro Internacional de Agricultura Tropical (CIAT), v. 22, n. 3, p. 2–15, 2000.

SANO, E. E. et al. Mapeamento semidetalhado do uso da terra do bioma cerrado. *Pesquisa Agropecuária Brasileira*, SciELO Brasil, v. 43, n. 1, p. 153–156, 2008.

WALTER, B. M. T. *Fitofisionomias do bioma Cerrado: síntese terminológica e relações florísticas.* Tese (Doutorado em Ecologia) — Universidade de Brasília, 2006.

WITTEN, I.; FRANK, E. *Data Mining: Practical Machine Learning Tools and Techniques With Java Implementations*. New York and San Mateo, CA: Morgan Kaufmann Publishers, Academic Press, 2000.