

Classificação de áreas algodoeiras utilizando séries temporais de imagens Modis

João Paulo Sampaio Werner¹
Stanley Robson de Medeiros Oliveira²
Júlio César Dalla Mora Esquerdo³

Resumo: O objetivo deste trabalho foi classificar áreas com o cultivo de algodão no estado de Mato Grosso. Foram utilizados dados de referência terrestre e uma série temporal de imagens de índices vegetativos do sensor Modis. A partir da série temporal de imagens, foram gerados perfis espectro-temporais e extraídas 11 métricas fenológicas em imagens de decomposição. A partir dessas métricas fenológicas, a técnica de indução de árvore de decisão foi aplicada para gerar regras de classificação que, posteriormente, foram utilizadas para separar as áreas com cultivo de algodão de outras coberturas vegetais. Os resultados encontrados, utilizando os índices de vegetação (NDVI e EVI) obtidos de séries temporais de imagens, demonstraram a capacidade dos modelos para discriminar áreas de algodão de outras culturas.

Palavras-chave: índice de vegetação, métricas fenológicas, algodão, mineração de dados.

¹Engenheiro-agrônomo, estudante de mestrado da Faculdade de Engenharia Agrícola da Universidade Estadual de Campinas (Unicamp), bolsista do CNPq.

²Cientista da Computação, PhD. em Ciência da Computação, pesquisador da Embrapa Informática Agropecuária, Campinas, SP.

³Engenheiro-agrônomo, doutor em Engenharia Agrícola, pesquisador da Embrapa Informática Agropecuária, Campinas, SP.

Introdução

Técnicas de sensoriamento remoto e classificação digital de imagem têm sido utilizadas para classificar áreas de cultivo agrícola. No entanto, diversas metodologias baseadas nessas técnicas utilizam variáveis espectrais obtidas apenas de uma ou poucas imagens em períodos específicos. Ocorre que alguns objetos na superfície terrestre podem ter um comportamento espectral similar ao da classe de interesse, podendo dificultar o processo de classificação do alvo pretendido. Outro agravante é que uma cultura pode se encontrar em diferentes estádios fenológicos e se apresentar com diferentes padrões espectrais, mesmo pertencendo a uma mesma classe de uso da terra. Esses desafios dificultam a elaboração de metodologias para classificar culturas agrícolas.

Uma das alternativas para minimizar esses problemas é incluir o domínio do tempo na abordagem, com o uso de séries temporais de imagens, uma vez que uma cultura agrícola segue um padrão quando seu perfil espectro-temporal é projetado.

O grande desafio é então automatizar o processo de identificação das principais culturas praticadas em escala regional. Uma das alternativas é o desenvolvimento de métodos sistemáticos que aprendam com os padrões espectro-temporais de culturas agrícolas, que são baseados nas informações extraídas em séries temporais de imagens.

A caracterização espectro-temporal das culturas agrícolas pode ser realizada usando os índices de vegetação, em especial o *Normalized Difference Vegetation Index* (NDVI) e o *Enhanced Vegetation Index* (EVI), que possuem alta correlação com a biomassa verde do dossel vegetal, e podem ser obtidos pelos produtos do sensor *Moderate Resolution Imaging Spectroradiometer* (Modis), que apesar de fornecer imagens com moderada resolução espacial, dispõe de excelente resolução temporal, em função de sua cobertura diária e amplo campo de visada.

Por outro lado, técnicas de mineração de dados têm demonstrado forte potencial em mapeamentos de culturas agrícolas, uma vez que permitem transformar as informações contidas em grandes volumes de dados em conhecimento e, portanto, são de grande aplicabilidade em estudos que utilizam séries temporais.

Este trabalho visa promover ganhos na automatização dos processos de identificação e classificação das áreas agrícolas com a cultura do algodão em Mato Grosso. O objetivo é gerar modelos eficientes baseados no método de indução por meio de árvore de decisão para classificação de áreas com cultivo de algodão, em escala regional, utilizando séries temporais de índices vegetativos do sensor orbital Modis.

Materiais e Métodos

A área de estudo compreende o estado de Mato Grosso, que está localizado na região Centro-Oeste do Brasil, com uma área de 903 mil km². O estado é responsável por 62% da área plantada de algodão no Brasil (ACOMPANHAMENTO DA SAFRA BRASILEIRA [DE] GRÃOS, 2016).

Como o trabalho é fundamentado pela abordagem espectro-temporal, foram utilizados subprodutos das imagens digitais multitemporais do sensor Modis, a bordo dos satélites Terra e Aqua, para extração de métricas fenológicas que descrevem o desenvolvimento das coberturas vegetais. As imagens que correspondem à área de estudo englobam quatro safras agrícolas, entre 2012/2013 e 2015/2016, com a finalidade de identificar a área ocupada pela cultura do algodoeiro em cada uma delas. Cada safra é composta por 46 composições máximas de 16 dias oriundas de ambos satélites, totalizando 184 imagens para cada índice vegetativo, com 250 m de resolução espacial, geradas a partir de passagens diárias do sensor Modis.

As imagens processadas do NDVI e EVI originadas do sensor Modis são disponibilizadas gratuitamente pelo Banco de Produtos Modis, mantido pela Embrapa Informática Agropecuária (ESQUERDO et al., 2011).

Como referência terrestre foram utilizados dados de localização geográfica, cedidos pela Embrapa Informática Agropecuária, das áreas de cultivo de algodão obtidas por trabalhos de campo, bem como outras culturas anuais praticadas na região, que serviram como amostras para obtenção da informação relativa ao padrão de perfil espectro-temporal médio das culturas.

A análise dos perfis temporais das áreas de referência foi realizada no programa *Timesat*, desenvolvido por Jönsson e Eklundh (2004). O uso do programa gerou métricas da fenologia do algodão a partir das imagens de índices de vegetativos, filtradas matematicamente por meio do filtro Savitzky-Golay (SAVITZKY; GOLAY, 1964). Este filtro serve para suavizar o perfil espectro-temporal,

baseia-se em uma janela móvel (tamanho 4), que utiliza ajuste linear de mínimos quadrados por meio de sucessivas equações polinomiais.

As métricas fenológicas permitem a discriminação de culturas agrícolas por meio da investigação dos diferentes efeitos causados pela sazonalidade, que se relacionam com o desenvolvimento vegetativo ao longo do tempo. A Figura 1 ilustra a série temporal do NDVI suavizada (em vermelho), resultado da aplicação de um de seus filtros disponíveis, e a série temporal original sem o uso de filtro (em azul), demonstrando também as onze métricas fenológicas que podem ser extraídas das séries temporais.

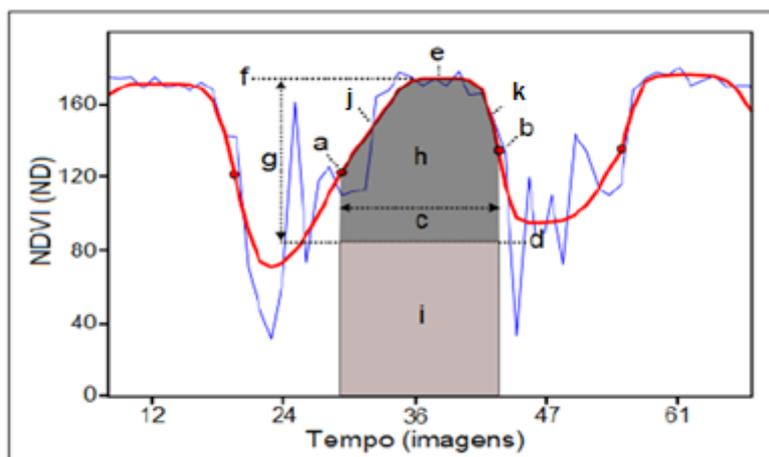


Figura 1. Métricas fenológicas sazonais extraídas pelo Timesat: (a) início do ciclo (b) final do ciclo (c) comprimento do ciclo (d) valor base (e) posição da metade do ciclo (f) pico de máximo (g) amplitude sazonal (h) pequena integral (produtividade primária) (i) grande integral (produtividade total, h + i) (j) taxa de crescimento e (k) taxa de senescência.

Fonte: adaptado de Eklundh e Jönsson (2015).

Um dos possíveis resultados de saída do *Timesat* são as imagens de decomposição, que representam cada uma das onze métricas fenológicas disponíveis, considerando os dois índices de vegetação. Essas imagens foram compiladas e executadas em rotinas específicas, escritas em linguagem *Interactive Data Language* (IDL), para extração de seus valores e conversão para formato de planilha, legível pelo software *Waikato Environment for Knowledge Analysis (Weka)* (WITTEN; FRANK, 2005). Na extração usou-se a localização geográfica presente nos dados de referência terrestre.

Foram gerados dois bancos de dados constituídos pelos valores das métricas fenológicas e sazonais derivadas das séries temporais de NDVI e EVI, respectivamente, extraídas no *Timesat*. As 11 métricas representam as variáveis independentes (atributos numéricos) e o atributo-meta contém duas classes: a) algodão (cultura de interesse); b) não algodão (todos os demais usos e coberturas da terra). Assim, os conjuntos de dados ficaram constituídos por 3.211 instâncias, para o conjunto do NDVI, e 2.890 instâncias para o conjunto do EVI, com 12 atributos em cada. No primeiro, havia 1.159 instâncias classificadas como algodão, enquanto no segundo, 1.167 instâncias.

As métricas fenológicas foram calculadas a partir da identificação das datas de início e final dos ciclos. Quando tais momentos não foram encontrados pelo *Timesat*, nenhuma outra métrica foi calculada. Isso explica a diferença no número de instâncias encontradas pelo NDVI e EVI, uma vez que as curvas desses dois índices podem apresentar comportamentos distintos para um mesmo pixel, possibilitando que as datas de início não fossem identificadas da mesma forma.

Em seguida, foi utilizado o algoritmo de indução por árvore de decisão J48, conhecido como C4.5, desenvolvido por Quinlan (1993). Para tornar o modelo mais simples e genérico, melhorar a taxa de acerto do classificador e diminuir o sobreajuste, foram testados diferentes números de objetos por folha para podar a árvore de decisão. Além disso, quatro diferentes métodos para seleção dos atributos foram testados, com o propósito de retirar os atributos com baixa correlação em relação às classes: a) sem seleção de atributos, onde ocorreu a utilização de todos atributos, caracterizando-se pela ausência de seleção; b) seleção de atributos baseado em correlação (CFS), que pesquisa o conjunto de atributos correlacionados com a classe e não correlacionados entre si; c) o método *InfoGain*, que avalia o valor de um atributo medindo o ganho de informação em relação à classe; d) a abordagem *Wrapper*, que ocorre conjuntamente com o algoritmo básico de aprendizagem, em que a validação cruzada é utilizada para estimar a precisão do esquema de aprendizagem para um conjunto de atributos (HAN et al., 2012).

O modelo de indução foi gerado por meio de validação cruzada (10 folds) e avaliado pelas medidas:

acurácia, coeficiente Kappa e precisão por classe, calculadas a partir da matriz de confusão (HAN et al., 2012).

Resultados e Discussão

Foram gerados diversos modelos por meio do algoritmo J48, considerando diferentes métodos de seleção de atributos. As Tabelas 1 e 2 apresentam, de uma maneira geral, as avaliações dos modelos para os índices de vegetação NDVI e EVI, respectivamente. Observa-se que as variáveis independentes, representadas pelas métricas fenológicas, conseguiram predizer as classes com um ótimo desempenho para os dois índices de vegetação. Para os classificadores baseados no NDVI, nota-se a superioridade do método Wrapper em relação ao demais, o qual conseguiu realizar a predição das classes, com alta acurácia, com apenas seis dos 11 atributos.

Tabela 1. Desempenho do classificador J48 para série temporal do NDVI.

Método de seleção	Acurácia	Kappa	Precisão por classe		Nº de regras	Nº obj por folha	Atributo Selecionado *
			Algodão	Não Algodão			
CFS	95,14%	0,89	0,95%	0,95%	12	6	i,h,e
Infogain	95,02%	0,89	0,94%	0,96%	12	8	h,i,c,f,g,b
Wrapper	96,08%	0,91	0,95%	0,97%	13	6	g,c,j,a,i,h

* Atributos selecionados: (a) início do ciclo; b) final do ciclo; c) comprimento do ciclo; d) valor base; e) posição da metade do ciclo; f) pico de máximo; g) amplitude sazonal; h) pequena integral (produtividade primária); i) grande integral (produtividade total); j) taxa de crescimento; e k) taxa de senescência.

Para os classificadores baseados no EVI, o modelo gerado por meio do método CFS, com apenas cinco atributos, apresentou resultados melhores, possivelmente pelo fato deste método ser mais eficiente para lidar com atributos numéricos.

Tabela 2. Desempenho do classificador J48 para série temporal do EVI.

Método de seleção	Acurácia	Kappa	Precisão por classe		Nº de regras	Nº obj por folha	Atributo Selecionado*
			Algodão	Não Algodão			
CFS	95,26%	0,9	0,95%	0,96%	8	6	c,i,h,f,e
Infogain	94,80%	0,89	0,94%	0,95%	10	6	h,i,c,g,f,b
Wrapper	94,91%	0,89	0,94%	0,95%	11	6	j,i,h,e

* Atributos selecionados: a) início do ciclo; b) final do ciclo; c) comprimento do ciclo; d) valor base; e) posição da metade do ciclo; f) pico de máximo; g) amplitude sazonal; h) pequena integral (produtividade primária); i) grande integral (produtividade total); j) taxa de crescimento; k) taxa de senescência.

As Tabelas 1 e 2 também apresentam os resultados para os diferentes valores de objetos por folha, variando de dois a 12. Para os modelos baseados no NDVI (Tabela 1), os melhores valores variaram entre seis e dez, enquanto para o EVI, o número seis apresentou os melhores resultados. Para ambos os conjuntos de dados, as métricas fenológicas: produtividade primária, produtividade total e comprimento do ciclo foram os atributos que apresentaram o maior ganho de informação.

Considerações Finais

Neste trabalho são apresentados classificadores para a discriminação de áreas com o cultivo de algodão, utilizando índices de vegetação NDVI e o EVI, obtidos a partir de séries temporais do sensor Modis. Os modelos preditivos gerados, com menos de 50% dos atributos, mostraram-se eficazes para classificar a cultura do algodão no estado de Mato Grosso. Como trabalho futuro, as regras geradas pelos modelos serão usadas para estimar e mapear a área plantada de algodão, considerando todas as quatro safras avaliadas (2012/2013 a 2015/2016) na área de estudo. Os resultados serão comparados com as estimativas oficiais, divulgadas por órgãos governamentais.

Referências

ACOMPANHAMENTO DA SAFRA BRASILEIRA [DE] GRÃOS: safra 2016/17: décimo levantamento, Brasília, DF, v. 4, n. 10, jul. 2016. 170 p. Disponível em: <http://www.conab.gov.br/OlalaCMS/uploads/arquivos/17_01_11_11_30_39_boletim_graos_janeiro_2017.pdf>. Acesso em: 5 fev. 2017.

EKLUNDH, L.; JÖNSSON, P. **Timesat 3.2 software manual**. Lund: Lund University; Malmö: Malmö University, 2015. 88 p.

ESQUERDO, J. C. D. M.; ANTUNES, J. F. G.; ANDRADE, J. C. de. Desenvolvimento do banco de produtos MODIS na base estadual brasileira. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 15., 2011, Curitiba. **Anais...** São José dos Campos: Inpe, 2011. p. 7596-7602. SBSR 2011.

HAN, J.; KAMBER, M.; PEI, J. **Data mining: concepts and techniques**. 3rd ed. Amsterdam: Morgan Kaufmann, 2012. 703 p.

JÖNSSON, P.; EKLUNDH, L. Timesat-a program for analyzing time-series of satellite sensor data. **Computers and Geosciences**, v. 30, n. 8, p. 833-845, Oct. 2004.

QUINLAN, J. R. **C4.5: programs for machine learning**. Amsterdam: Morgan Kaufmann, 1993. 302 p. il. (The Morgan Kaufmann series in machine learning).

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1627-1639, 1964.

WITTEN, I. H.; FRANK, E. **Data mining: practical machine learning tools and techniques**. 2nd ed. Amsterdam: Elsevier, 2005. 525 p. il. (Morgan Kaufmann series in data management systems).