

COMPARAÇÃO ENTRE OS CLASSIFICADORES SUPPORT VECTOR MACHINE E RANDOM FOREST NA IDENTIFICAÇÃO DE OBJETOS PRESENTES EM CENAS CAFEEIRAS OBTIDAS POR VANT

Samuel Lima Ferreira de Araujo¹ e Thomé Simpliciano Almeida²

¹IFSULDEMINAS - Campus Passos, samuel.araujo@alunos.ifsuldeminas.edu.br; ²IFSULDEMINAS - Campus Passos, thome.almeida@ifsuldeminas.edu.br

RESUMO

O monitoramento das condições de uma determinada cultura agrícola é essencial para tomadas de decisão que visem a melhoria do produto na agricultura. Com o desenvolvimento dos VANTs e dos sensores, o gerenciamento dos problemas relacionados a lavoura se tornou mais acessível. Com isso, o trabalho apresenta uma comparação dos classificadores *Support Vector Machine* (SVM) e *Random Forest* (RF) na identificação de objetos presentes em uma plantação de café. A partir dos objetos obtidos com o método de segmentação “crescimento de regiões”, foram extraídos atributos estatísticos para o treinamento dos classificadores. Como resultado foi observado um melhor desempenho do SVM com a acurácia de 98,6% e o índice *Kappa* de 0,978, por sua vez o classificador RF obteve a acurácia de 96,7% e o índice *Kappa* de 0,947.

Palavras-chave — OBIA, agricultura de precisão, sensoriamento remoto, aprendizado de máquina.

ABSTRACT

The monitoring of a determined agricultural culture is essential to the decision making of product improvement in agriculture. The development of UAVs and sensors allowed the management of crop problems to become more accessible. Thus, the work presents a comparison of Support Vector Machine (SVM) and Random Forest (RF) classifiers in the identifying die back's in coffee crops. From the objects obtained through the segmentation method “region growing”, statistical attributes were extracted for the training step. The SVM classifier obtained the best result with the accuracy of 98.6% and kappa index of 0.978, the RF classifier obtained the accuracy of 96.7% and the kappa index of 0.947.

Key words — OBIA, precision agriculture, remote sensing, machine learning.

1. INTRODUÇÃO

Um tema que se tornou importante na agricultura de precisão é o monitoramento de culturas agrícolas através de imagens aéreas. O desenvolvimento dos veículos aéreos não tripulados

(VANTs) acoplados a câmeras digitais proporcionam à essa plataforma utilidade para o sensoriamento remoto [1]. Com isso, informações sobre o estado atual do plantio e identificação de problemas que limitam a produtividade são apresentados com o uso do sensoriamento remoto na agricultura [2]. O gerenciamento das lavouras, identificação de pragas e condições de déficit hídrico e nutricional, são importantes em aplicações da agricultura de precisão, como exemplo, estimativa de produção, aplicação de herbicidas, ocasionando em reduções de custos da produção e impactos ambientais [1].

Um novo desafio na classificação de imagens de sensoriamento remoto deve-se ao fato do acesso de dados de alta resolução [3]. Com isso, o método *object-based image analysis* (OBIA) se tornou um substituto do tradicional método baseado em pixels. No contexto da classificação, o aprendizado de máquina provou ser útil para várias aplicações do sistema terrestre e detecção de doenças em lavouras [4].

A proposta deste trabalho é a identificação de objetos presentes em plantações cafeeiras utilizando imagens obtidas por câmera em um VANT. Para a identificação das classes, foram utilizados os classificadores *Support Vector Machine* (SVM) e *Random Forest* (RF).

2. MATERIAIS E MÉTODOS

Inicialmente, foi selecionada a região cafeeira experimental do IFSULDEMINAS - Campus Machado - MG para realizar as obtenções das imagens. Para o imageamento, foi utilizada a câmera presente no *Phantom 4* (capaz de adquirir imagens do espectro RGB com imagens de 12 megapixels), com vôos realizados a uma altura de 100 metros.

A segmentação deu-se pelo método baseado em regiões, utilizou-se a ferramenta “Crescimento de Regiões Médio” do software *TerraView*, os parâmetros “tamanho mínimo de segmentos” alterados sob o intervalo de [50, 60, 70, ..., 100], e o “limiar de similaridade”, com variação de [0,001, 0,002, 0,003, ..., 0,03]. As bandas RGB e o índice de vegetação *Excess Green* (ExG) (Fórmula 1) foram utilizados na segmentação como parâmetro de entrada. Com isso, permitiu-se a escolha dos valores que garantisse a melhor divisão dos objetos presentes na cena.

$$ExG = 2 * \frac{G}{R + G + B} - \frac{R}{R + G + B} - \frac{B}{R + G + B}$$

Onde:

G = valores do canal verde

R = valores do canal vermelho

B = valores do canal azul

Após a etapa de segmentação, foi realizada a extração dos atributos de cada região. Para cada região, foram extraídos os atributos: média, mediana, valor mínimo e máximo, variância e amplitude. Os alvos selecionados durante a amostragem foram reconhecidos utilizando informações obtidas em campo e através da identificação visual nas imagens. Essas regiões foram divididas em dois propósitos, sendo eles para treinamento (75%) e para validação (35%) do classificador.

As implementações do SVM e do RF estão presentes no módulo *Scikit-learn*, para o aperfeiçoamento dos parâmetros dos classificadores foi utilizado o método *Grid Search*, que realiza uma busca exaustiva dado um conjunto de valores para os parâmetros que garanta um melhor desempenho na etapa de classificação. Como parâmetros para a função *Grid Search* do classificador SVM foram passados os *kernels* polinomial, RBF e linear, com o parâmetro *gamma* variando de $[1 \times 10^{-5}, 1 \times 10^{-4}, \dots, 1 \times 10^{-2}]$ e o parâmetro C variando de $[1 \times 10^{-2}, 1 \times 10^{-1}, \dots, 1 \times 10^3]$. Já para o classificador RF o ajuste foi realizado sobre dois parâmetros, para o parâmetro “*n estimators*” foi passado os valores 1×10^1 , 1×10^2 e 1×10^3 e para o parâmetro “*max features*” foram passados os valores 1, 3 e 5.

Para a análise das classificações, foram empregues a matriz de confusão, o índice *Kappa*, *mean absolute error* (MAE), *root mean squared error* (RMSE) e *mean squared error* (MSE).

3. RESULTADOS E DISCUSSÃO

O resultado obtido pelo método *Grid Search* aplicado foi o seguinte modelo: *kernel* polinomial, 1×10^{-2} para o parâmetro C , 0 para o parâmetro r , 3 para o parâmetro d e 1×10^{-4} ao parâmetro γ . Os resultados das análises de erro, para o modelo indicado, são apresentados na Tabela 1.

Método	Support Vector Machine
<i>Kappa</i>	0,978
MAE	0,019
MSE	0,031
RMSE	0,176
<i>Accuracy</i>	0,986

Tabela 1. Valores da avaliação do classificador *Support Vector Machine*.

A Tabela 2 apresenta a matriz de confusão do classificador SVM.

	Cafeeiro	Seca dos ponteiros	Solo
Cafeeiro	216	2	1
Seca dos ponteiros	1	84	1
Solo	2	0	211

Tabela 2. Matriz de confusão do classificador *Support Vector Machine*.

Já para o classificador RF os melhores resultados foram obtidos com os parâmetros, valor 3 para o “*max features*”, 2 para o “*min samples split*”, 100 para o “*n estimators*” e nulo para o parâmetro “*max depths*”. A Tabela 3 apresenta as análises estatísticas obtidas com este modelo.

Método	Support Vector Machine
<i>Kappa</i>	0,947
MAE	0,035
MSE	0,039
RMSE	0,196
<i>Accuracy</i>	0,967

Tabela 3. Valores da avaliação do classificador *Random Forest*.

A Tabela 4 apresenta a matriz de confusão do classificador RF.

	Cafeeiro	Seca dos ponteiros	Solo
Cafeeiro	217	5	0
Seca dos ponteiros	1	76	5
Solo	1	5	208

Tabela 4. Matriz de confusão do classificador *Random Forest*.

A partir do treinamento dos modelos foram geradas imagens temáticas com as classes cafeeiro (verde), solo exposto (azul) e vermelho (seca dos ponteiros). A Figura 1 demonstra a imagem classificada pelo modelo do SVM.

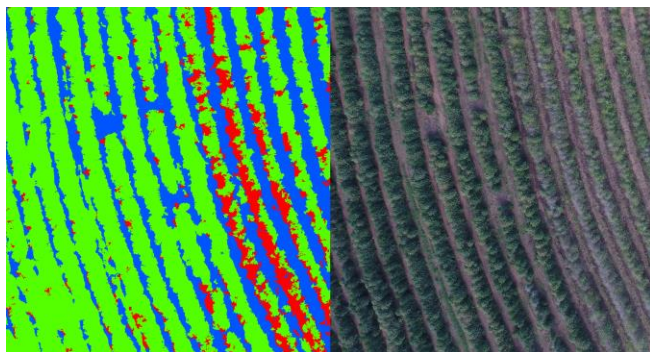


Figura 1. Imagem classificada pelo modelo SVM.

A Figura 2 apresenta a imagem classificada pelo modelo RF. Observa-se que houve mais predições para a classe “seca dos ponteiros”.

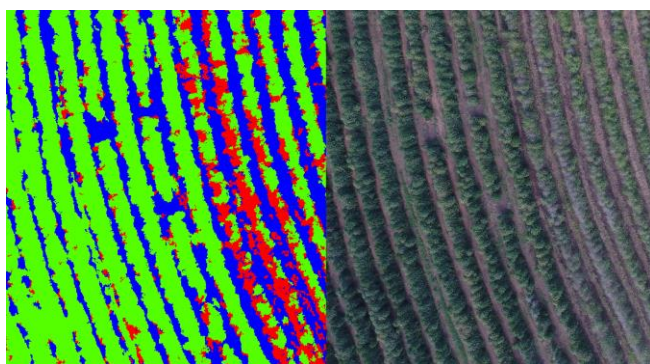


Figura 2. Imagem classificada pelo modelo RF.

O resultado obtido pelas classificações dos dois modelos apresentaram índices muito satisfatórios. Em um trabalho similar [5], onde o objetivo era o mapeamento de lavoura de café, a classificação do algoritmo SVM apresentou a acurácia 85,33% e o índice *Kappa* em 0,80. Já o classificador RF apresentou a acurácia de 77,66% e o índice *Kappa* 0,80. Neste trabalho, observa-se um melhor resultado sobre os dois índices, com o classificador SVM apresentando a acurácia de 98,6% e índice *Kappa* de 0,978 e o classificador RF apresentando a acurácia de 96,7% e índice *Kappa* de 0,947, apresentando um grau de concordância muito bom para os dois classificadores [6] (valores entre 0,81 e 1,0 para o índice *Kappa*).

5. CONCLUSÕES

Neste trabalho foram comparados dois métodos de classificação, o classificador *Random Forest* e o *Support Vector Machine*. Utilizou-se na segmentação o método crescimento de regiões e realizou-se a extração de atributos das regiões reconhecidas. Para uma melhor escolha dos parâmetros foi utilizado o método *Grid Search*, passando um conjunto de valores para ser analisado.

Os valores obtidos na análise dos classificadores mostram que os dois métodos se apresentaram adequados para classificação. Sendo que o método *Support Vector Machine* apresentou melhores valores do que o método *Random Forest* em todos os métodos de avaliação utilizados.

6. REFERÊNCIAS

- [1] Pajares, G. “Overview and Current Status of Remote Sensing Applications Based on Unmanned Aerial Vehicles (UAVs)”. *Photogrammetric Engineering & Remote Sensing*, v. 81, n. 4, p. 281-329, 2015.
- [2] Pinter, P.J.Jr. et al. Remote sensing for crop management. *Photogrammetric Engineering & Remote Sensing*, v. 69, n. 6, p. 647-664, 2003.
- [3] Ma, L. et al. A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 130, p. 277-293, 2017.
- [4] Lary, D.J. et al. Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, v. 7, n. 1, p. 3-10, 2016.
- [5] Souza, C.G. et al. Machine learning algorithms and variable of remote sensing for coffee cropping mapping. *Boletim de Ciências Geodésicas*, v. 22, n. 4, p. 751-773, 2016.
- [6] Galparsoro, L.U. e Fernández, S.P. Medidas de concordancia: el índice *Kappa*. *CaD Aten Primaria*, Coruña, n.6, p. 169-171, 1999. Disponível em: <http://www.fisterra.com/mbe/investiga/kappa/kappa.htm>. Acesso em: 01 out. 2018.