

Classificação por árvores de decisão: Avaliação de acurácia com e sem a pré-seleção de atributos

Camila Souza dos Anjos Lacerda¹
Cláudia Maria de Almeida²
Lênio Soares Galvão²

¹ Instituto de Estudos Avançados - IEAv/FAB
Trevo Cel. Av. José A. A. do Amarante, 01- 12228-001 - São José dos Campos - SP, Brasil
camilasa@ieav.cta.br

² Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12227-010 - São José dos Campos - SP, Brasil
{almeida, lenio}@dsr.inpe.br

Abstract. The use of decision trees for image classification has grown rapidly in recent years, since the reported developments are promising. The current work uses the C4.5 and Random Forest methods for selecting statistical and customized attributes derived from a WorldView-2 (WV-2) image, meant to effectively separate the classes of interest. The attributes comprising the bands extracted from transforms such as Principal Component Analysis and Minimum Noise Fraction in addition to vegetation indices and attributes based on band ratios were used in the stage of feature selection. Two datasets have been employed in the herein described experiments: one consisting of a WV-2 scene with 42 image-derived attributes, and the second one containing the same WV-2 scene and 28 pre-selected meaningful attributes. Comparisons between the two datasets showed that for both methods (C4.5 and RF) the use of statistical attributes plus those derived from image transforms and arithmetical operations increases the agreement indices accuracy. Both classifiers work as data miners, identifying among a large set of attributes those able to discriminate the concerned classes. The results of this article comply with the peer-reviewed literature, for they demonstrate that the classification integrating a greater number of input attributes (without feature selection) attain a significantly superior accuracy. In sum, decision tree classifiers have the capacity to deal with weak explanatory variables, and hence, it is not possible to assess their importance based on individual factors alone.

Palavras-chave: decision trees, feature selection, data mining, árvores de decisão, seleção de atributos, mineração de dados.

1. Introdução

A constante evolução dos sensores tem motivado pesquisadores a desenvolverem novos métodos de classificação, visando aproveitar a riqueza de conteúdo e ao mesmo tempo, atenuar as dificuldades impostas por sua elevada dimensionalidade (Hardin e Hardin, 2013).

Além do uso de novos métodos, o uso da imagem acrescido de atributos derivados da imagem original aumenta a acurácia da classificação (Tong et al, 2014). Neste sentido, as imagens de entrada sofrem transformações ou passam por operações aritméticas para gerar novos atributos, e assim, poder separar de forma mais efetiva as classes de interesse.

É comum utilizar transformações, métricas de textura, índices de vegetação, índices de água, além de atributos aritméticos customizados com base em operações entre bandas na etapa de extração de atributos dessas imagens.

O interesse no uso de árvores de decisão para a classificação de imagens tem crescido rapidamente nos últimos anos, visto que os desenvolvimentos relatados são promissores. De acordo com Tso e Mather (2009), a categorização de dados utilizando mecanismos de separação hierárquica tem sido amplamente utilizada.

Segundo Han et al. (2012), os algoritmos de árvores de decisão foram originalmente planejados para classificação, nos quais uma árvore de decisão é construída em uma estrutura de fluxograma (diagrama de fluxo). Cada nó interno indica um teste de um atributo. Cada ramo corresponde a um resultado do teste, e cada nó externo, também chamado de folha,

denota uma predição de classe. Assim, o resultado do classificador é encontrado utilizando uma sequência de decisões, em que um conjunto de regras é implementado, e estas determinam o caminho a ser seguido, começando do nó-raiz e finalizando no nó-terminal (folha), que representa a classe para a qual o objeto foi classificado (Tso e Mather, 2009). O problema da construção de uma árvore de decisão é encontrar os melhores atributos para discriminar as classes e dividir de forma recursiva os dados em subconjuntos (Wang e Li, 2008).

O presente trabalho utilizará dois conjuntos de dados, sendo um deles composto por uma cena Worldview-2 (WV-2) e 42 atributos derivados da mesma cena, e o outro conjunto contendo a mesma cena WV-2 e 28 atributos significativos pré-selecionados.

A classificação de imagens por árvores de decisão por meio de dois métodos diferentes C4.5 e Floresta Randômica (*Random Forest - RF*) analisa o desempenho das classificações em dois diferentes níveis de legenda, sendo um deles com 11 classes de cobertura, e o outro, mais detalhado, com a presença de todos os materiais visualmente identificados em campo, resultando em 38 classes de cobertura do solo urbano. Por fim se verifica se os dois conjuntos de dados (completo e pré-selecionado) produzem resultados de classificação em ambos os níveis de legenda significativamente distintos.

2. Material e Métodos

2.1 Material e Área de Estudo

A área de estudo do presente trabalho corresponde a um transecto situado no Campus da Universidade Estadual de Campinas – UNICAMP, localizada na cidade de Campinas, interior do estado de São Paulo. Esta área foi escolhida devido à diversidade de materiais presentes na sua superfície e pela facilidade de acesso para as visitas de campo.

As imagens WorldView-2 utilizadas no presente trabalho foram adquiridas em 21 de junho de 2010 e 02 de julho de 2010. As duas cenas WV-2 foram empregadas em conjunto somente para gerar o modelo digital de superfície (MDS) a ser usado no procedimento de ortorretificação. A partir de então, somente a cena de 02 de julho de 2010 foi utilizada, em virtude de a mesma apresentar a menor angulação off-nadir, correspondente a $6,75^\circ$, além de $38,37^\circ$ de elevação solar e 0% de cobertura de nuvem.

2.2 Metodologia de Trabalho

Inicialmente, converteram-se as imagens WV-2 para radiância, utilizando a ferramenta "Calibração Radiométrica" (*Radiometric Calibration*) do ENVI 5.0. Em seguida, as imagens foram transformadas para reflectância de superfície por meio do aplicativo "Análise Rápida de Linha de Visão Atmosférica de Hipercubos Espectrais" (*Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes - FLAASH*). O FLAASH realiza a correção atmosférica das cenas e apresenta os dados resultantes na grandeza física de reflectância de superfície.

Após a correção atmosférica, as bandas multiespectrais WV-2 com 2 m de resolução espacial foram fusionadas com a banda pancromática WV-2 com 0,5 m de resolução espacial. O método Gram-Schmidt foi aplicado utilizando o *software* ENVI 5.0. Tal método foi escolhido por não impor limite ao número de bandas de entrada e também por apresentar alto grau de similaridade espectral entre as bandas multiespectrais originais e as transformadas. As oito bandas WV-2, após a fusão dos dados, passaram a ter 0,5 m de tamanho de pixel.

Foi realizado um trabalho de campo para obtenção das coordenadas utilizadas na ortorretificação da imagem. Foram coletados 17 pontos precisos em campo. A estação fixa utilizada neste trabalho foi a SPC1/ Campinas-UNICAMP, que faz parte da Rede Brasileira de Monitoramento Contínuo dos Sistemas GNSS (RBMC).

Para a ortorretificação foi utilizada a ferramenta de "Dispositivo de Ortorretificação" (*OrthoEngine*) do PCI *Geomatics*. Para tal finalidade, foi necessário previamente elaborar o

MDS utilizando as duas cenas WV-2 (21/06/2010 e 02/07/2010). Somente após a elaboração do MDS, é que foi feita a ortorretificação apenas para a imagem com o menor ângulo de visada off-nadir (a cena de 02 de julho de 2010).

Foram gerados também o modelo digital de terreno (MDT) e o modelo digital de altura (MDA). O MDT, que representa a superfície do terreno por meio de uma grade regular de cotas altimétricas, foi gerado no *software* QGIS 2.8.3 por meio de uma interpolação entre as coordenadas dos pontos de controle obtidos em campo. O atributo MDA, também chamado por alguns autores de modelo digital de superfície normalizado (MDSn), é resultante da subtração entre MDS e MDT.

Todos os dados geográficos utilizados no presente trabalho foram referenciados à Projeção UTM (Universal Transversa de Mercator) no Fuso 23 e Zona K, e datum horizontal WGS-84 (Sistema Geodésico Global 1984 ou *World Geodetic System* 1984).

A extração de atributos (*Feature Extraction*) foi realizada com o objetivo de extrair importantes informações do conjunto de dados. Para tal, foram criados os atributos correspondentes às bandas extraídas da Análise por Componentes Principais (*Principal Component Analysis* - PCA) e Fração Mínima de Ruído (*Minimum Noise Fraction* - MNF), que foram geradas a partir das respectivas ferramentas no ENVI 5.0. Além das bandas da PCA e MNF, foi adicionado como atributo o MDA.

Outros atributos customizados (AC) foram gerados por operações aritméticas entre as faixas espectrais por meio da ferramenta Matemática de Bandas (*BandMath*) do ENVI 5.0. Tais operações matemáticas têm como resultado uma nova matriz de dados, ou seja, uma nova imagem customizada completamente distinta das imagens de entrada.

Os índices de vegetação também podem ser considerados atributos customizados, pois também são gerados por operações aritméticas entre as faixas espectrais. Entre os atributos customizados gerados nesta pesquisa, estão os índices NDVI (*Normalized Difference Vegetation Index* ou Índice de Vegetação por Diferença Normalizada) (Rouse et al., 1974) e SAVI (*Soil-Adjusted Vegetation Index* ou Índice de Vegetação Ajustado ao Solo) (Huete, 1988), que foram incluídos com seus nomes próprios por serem amplamente conhecidos na literatura com esta terminologia. Outros 23 atributos customizados derivados da imagem WV-2 foram adicionados como atributos. A formulação matemática de cada um deles, juntamente com suas referências bibliográficas estão descritas em Anjos (2016).

Visando particionar a imagem em segmentos homogêneos, a imagem foi segmentada por meio da ferramenta Segmentação Multirresolução do eCognition 8.7 com os seguintes parâmetros: Escala = 30, Forma = 0,8, Compacidade = 0,9 e Peso = 1 para todas as bandas. Após a segmentação, foi executado o ajuste por diferença espectral (*Spectral Difference*) = 500. Este algoritmo é projetado para refinar os resultados da segmentação, por meio da similaridade espectral entre os segmentos vizinhos (Definiens, 2007).

Segmentos amostrais para cada uma das classes nos dois diferentes níveis de detalhamento (Nível 1 com 11 classes e Nível 2 com 38 classes) foram coletados no eCognition 8.7 e exportados no formato CSV (*Comma Separated Values*). O nivelamento hierárquico da legenda de classificação foi adaptado de Herold et al. (2004) com classes devidamente adequadas à cobertura do solo na área de estudo.

Tais amostras no formato CSV foram posteriormente importadas pelo *software* livre WEKA (*Waikato Environment for Knowledge Analysis* ou Ambiente Waikato para Análise do Conhecimento), versão 3.7.12, para serem utilizadas como dados de entrada para a classificação dos dados.

Na intenção de formar um novo conjunto de dados, alguns atributos considerados menos significativos foram excluídos do processo por meio da pré-seleção de atributos. Assim, foram excluídas as oito bandas da PCA, visto que a transformação MNF é um procedimento estatístico que concilia tanto os procedimentos de segregação da componente ruído, quanto

também de redução da dimensionalidade dos dados sem perda de informação (Green et al., 1988). Também foram excluídas as seis últimas bandas da MNF, dado que as componentes são ordenadas conforme a qualidade da imagem expressa pelo nível de ruído presente na cena.

Após a coleta e exportação das amostras no formato CSV e a seleção de atributos no segundo conjunto de dados, foram aplicados os métodos de classificação por árvore de decisão C4.5 e RF no *software* WEKA para os dois conjuntos de dados. Maiores informações sobre os classificadores C4.5 e RF podem ser obtidos respectivamente em Quinlan (1993) e Breiman (2001).

Para gerar as árvores RF, indicou-se o número de 10 árvores para cada uma das quatro execuções do algoritmo. Assim, foram originadas 10 árvores para cada conjunto de dados em cada nível de legenda (detalhamento), totalizando assim 40 árvores RF e mais 4 árvores C4.5. Todas as 44 árvores (para os dois conjuntos de dados WV-2, com e sem pré-seleção de atributos, em ambos os níveis de legenda) foram implementadas no eCognition 8.7 por meio da ferramenta “Atribuir Classe” (*Assign Class*), em que o atributo escolhido para cada nó da árvore é selecionado e é atribuído a ele o respectivo “Limiar de Condição” (*Threshold Condition*) determinado pela árvore.

Após a implementação de cada árvore, todas elas foram exportadas no formato *shapefile*, de forma que a informação do nome da classe associada a cada segmento estivesse presente no arquivo de formato DBF (*Database File*). Utilizando a informação do nome da classe de cada segmento, um programa em Visual Basic executado no Excel foi desenvolvido para realizar a contagem de votos (*voting*) das árvores RF. Dada a existência de 10 árvores RF para cada experimento de classificação, a classificação final RF é dada em função da votação da maioria de votos pertencentes a todas as árvores para um determinado objeto, ou seja, para cada segmento será atribuída a classe expressa pela maioria dos votos das 10 classificações realizadas por RF para cada experimento.

Assim, um novo arquivo DBF foi gerado a partir da votação, apresentando um novo *shapefile* com o resultado final do método para cada conjunto de dados em cada nível de legenda. Por fim, as oito classificações do presente trabalho (quatro classificações finais RF e quatro classificações C4.5) foram visualizadas no *software* QGIS 2.8.3.

Obtiveram-se então oito classificações, sendo quatro delas geradas pelo conjunto de dados WV-2 com um conjunto total de 42 atributos para os métodos C4.5 e RF nos dois níveis de detalhamento, e as outras quatro, para o conjunto WV-2 com pré-seleção de 28 atributos com os classificadores C4.5 e RF, também nos dois níveis de detalhamento.

A validação dos resultados foi verificada junto ao mapa de referência elaborado com o auxílio de fotointérpretes especialistas (Anjos, 2016). Foram gerados oito conjuntos de coordenadas aleatórias dentro do limite da cena no *software* Minitab 17, sendo conjuntos amostrais independentes para cada uma das oito classificações. Cada conjunto de pontos foi importado para o QGIS 2.8.3, gerando oito arquivos de pontos aleatórios no formato *shapefile*. Para cada *shapefile* de pontos, foi realizado o cruzamento espacial entre a localização dos pontos aleatórios nos dados de referência e em cada classificação. Por meio dos dados adicionados nas tabelas de atributos do *shapefile* de pontos, foram geradas as matrizes de confusão, através das quais foram calculados a exatidão global e o índice Kappa para cada classificação. Tal procedimento foi repetido oito vezes, isto é, para cada classificação final referente a cada conjunto de dados, classificador e nível de detalhamento.

Por fim, os testes de hipótese Z com nível de significância de 5% foram realizados entre os resultados obtidos pelo mesmo classificador, no mesmo nível de detalhamento, para os diferentes conjuntos de dados (com 42 atributos e pré-selecionado com 28 atributos), a fim de determinar se as diferenças de acurácia entre as distintas classificações são estatisticamente significantes.

3. Resultados e Discussão

A legenda de classes do Nível de Legenda 1 é apresentada na Figura 1. As classificações obtidas para o citado nível de detalhamento são apresentadas na Figura 2.



Figura 1. Legenda para as classificações do Nível de Legenda 1.

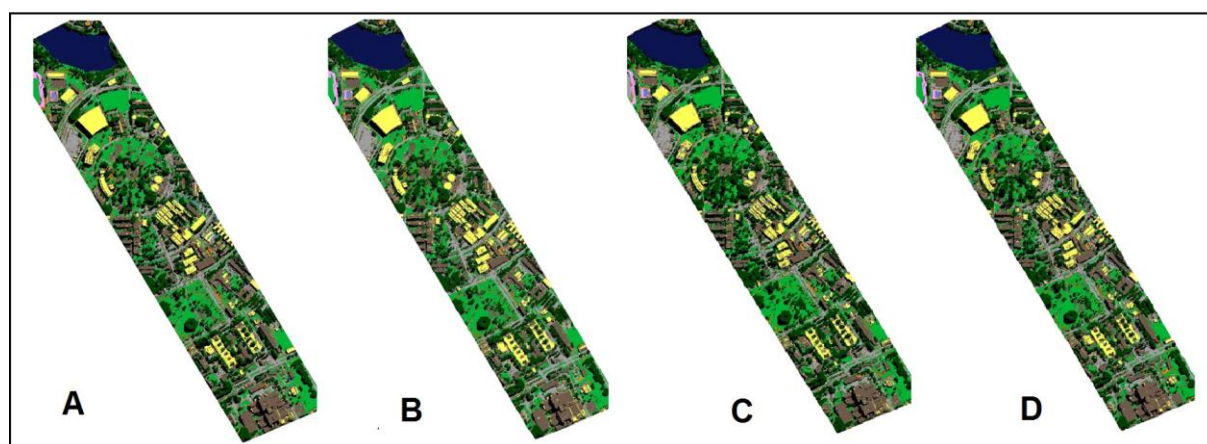


Figura 2. Classificações para o Nível de Legenda 1. (A) Classificador C4.5 com atributos pré-selecionados. (B) Classificador RF com atributos pré-selecionados. (C) Classificador C4.5 com o conjunto completo de atributos. (D) Classificador RF com o conjunto completo de atributos.



Figura 3. Legenda para as classificações do Nível de Legenda 2.

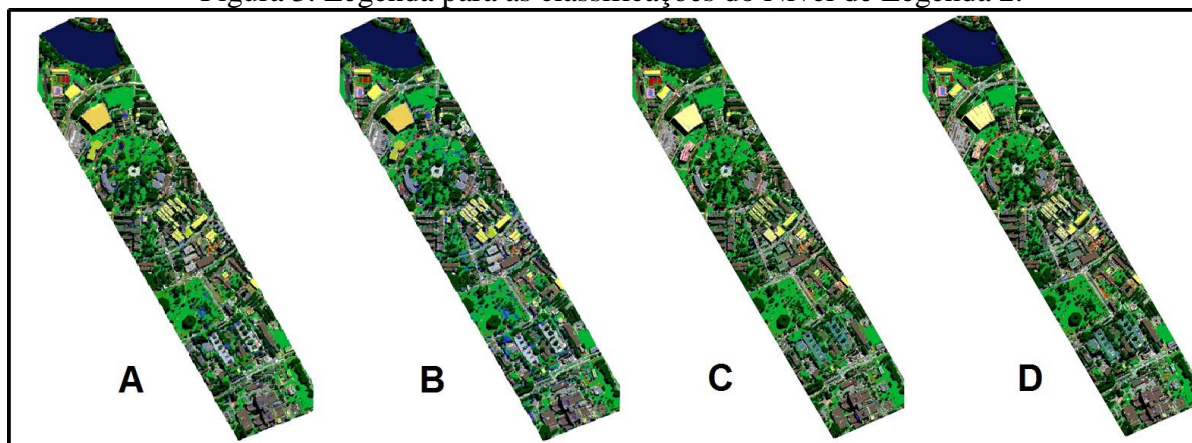


Figura 4. Classificações para o Nível de Legenda 2. (A) Classificador C4.5 com atributos pré-selecionados. (B) Classificador RF com atributos pré-selecionados. (C) Classificador C4.5 com o conjunto completo de atributos. (D) Classificador RF com o conjunto completo de atributos.

Visto que o enfoque do presente artigo é a pré-seleção de atributos, os resultados numéricos serão apresentados separadamente por conjunto de dados. A Tabela 1 apresenta os valores de Exatidão Global e índice Kappa das classificações realizadas com todos os atributos e considerando ambos os níveis de detalhamento.

Tabela 1. Exatidão Global e índice Kappa das classificações realizadas com o conjunto completo de atributos.

Nível de detalhamento	Classificador	Exatidão Global (%)	Índice Kappa
1	C4.5	85,60	0,8333
1	RF	86,94	0,8496
2	C4.5	80,17	0,7807
2	RF	76,28	0,7430

A Tabela 2 apresenta os valores de Exatidão Global e índice Kappa das classificações realizadas para o conjunto pré-selecionado de atributos.

Tabela 2. Exatidão Global e índice Kappa das classificações realizadas com pré-seleção de atributos.

Nível de detalhamento	Classificador	Exatidão Global (%)	Índice Kappa
1	C4.5	73,77	0,6937
1	RF	78,91	0,7541
2	C4.5	70,07	0,6712
2	RF	74,94	0,7215

Por fim, quatro testes de hipóteses foram realizados. O teste de hipótese Z com nível de significância de 5% foi realizado entre pares de classificadores de um mesmo método e mesmo nível de detalhamento com distintos conjuntos de dados. Eles examinaram comparativamente as classificações, visando averiguar se os índices Kappa das classificações comparadas são ou não significativamente diferentes. As hipóteses testadas e os resultados são apresentados resumidamente na sequência:

- Teste 1 - Comparação entre os conjuntos de dados completo e com pré-seleção de atributos para o classificador C4.5 no Nível de Legenda 1. Resultado: o índice Kappa do conjunto de dados completo é significativamente maior que o índice Kappa do conjunto de dados pré-selecionado.
- Teste 2 - Comparação entre os conjuntos de dados completo e com pré-seleção de atributos para o classificador RF no Nível de Legenda 1. Resultado: o índice Kappa do conjunto de dados completo é significativamente maior que o índice Kappa do conjunto de dados pré-selecionado.
- Teste 3 - Comparação entre os conjuntos de dados completo e com pré-seleção de atributos para o classificador C4.5 no Nível de Legenda 2. Resultado: o índice Kappa do conjunto de dados completo é significativamente maior que o índice Kappa do conjunto de dados pré-selecionado.
- Teste 4 - Comparação entre os conjuntos de dados completo e com pré-seleção de atributos para o classificador RF no Nível de Legenda 2. Resultado: Os Kappas não são significativamente diferentes.

Os atributos excluídos durante o processo de pré-seleção de atributos foram atributos considerados redundantes, fracos e/ou com grande ocorrência de ruídos. Entretanto, as comparações entre os dois conjuntos de dados demonstram que, para ambos os métodos (C4.5 e RF), as classificações com o maior número de atributos apresentam resultados significativamente superiores, exceto para o classificador RF no Nível de Legenda 2, em que os resultados, apesar de superiores no conjunto completo, não foram considerados significativamente diferentes.

Os dois classificadores escolhidos (C4.5 e RF) trabalham como mineradores de dados, identificando, em meio a um amplo conjunto de entrada, quais atributos são necessários para determinar a separação entre as classes.

Os resultados do presente artigo corroboram o trabalho de Breiman (2001), que afirma que é impossível compreender individualmente cada uma das variáveis de uma Floresta Aleatória. Logo, não é possível verificar sua importância baseando-se em fatores individuais. Isto vem de encontro aos achados de Walton (2008), para quem o classificador RF possui a capacidade de lidar com variáveis explicativas fracas.

4. Conclusões

Foram realizadas quatro diferentes classificações para a cena WV-2 e seus 42 atributos e outras quatro classificações para o conjunto composto pela cena WV-2 com 28 atributos pré-selecionados.

Os métodos utilizados (C4.5 e RF) permitem a inserção de um grande número de variáveis no algoritmo sem a preocupação, *a priori*, de se saber se estas serão explicativas ou não. De acordo com Breiman (2001), em relação ao RF, a importância de uma variável é estimada posteriormente por mudanças no erro de predição durante a execução do algoritmo.

Breiman e Cutler (2004) afirmam que uma das características do RF é a capacidade de lidar com milhares de variáveis de entrada sem eliminação prévia de variáveis.

Os resultados do presente trabalho, juntamente com a informação encontrada na literatura indexada, confirmam que, para ambos os métodos avaliados, os resultados obtidos por conjunto de dados previamente selecionados não são superiores aos resultados dos conjuntos de dados com maior quantidade de atributos. Tal fato se explica pelo motivo de que os métodos que utilizam árvores de decisão trabalham como mineradores de dados, selecionando os melhores atributos e lidando com grande quantidade de variáveis, além da capacidade de trabalhar com variáveis explicativamente fracas.

Sugere-se, para trabalhos futuros, testar-se a eficácia de algoritmos específicos para a pré-seleção de atributos, como os desenvolvidos por Kohavi e John (1997); Hall (1998) e Guyon et al. (2002).

Referências Bibliográficas

- Anjos, C. S. **Classificação de Áreas Urbanas com Imagens Multiespectrais e Hiperespectrais Utilizando Métodos Não-Paramétricos**. 2016. 336p. Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2016.
- Breiman, L. Random Forests. **Machine Learning**, V. 45, N. 1, P. 5-32, 2001.
- Breiman, L.; Cutler, A. **Random Forests**. 2004. Disponível em: <<https://www.stat.berkeley.edu/~breiman/RandomForests>>. Acesso em: 10 jan. 2016.
- Definiens. **Definiens Developer 7**: reference book. Munique: Definiens AG, 2007. 195 p. Disponível em: <<http://www.ecognition.cc/download/ReferenceBook.pdf>>. Acesso em: 20 jan. 2016.
- Green, a. A.; berman, m.; switzer, p.; craig, m. D., "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," **IEEE Transactions on Geoscience and Remote Sensing**, v. 26, n. 1, p. 65-74, 1988. DOI 10.1109/36.3001. Disponível em: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=3001&isnumber=170>>. Acesso em: 20 Jan. 2016.
- Guyon, I; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. **Machine Learning**. v.46, p.389-422, 2002. Disponível em: <<http://www.clopinet.com/isabelle/Papers/geneselect.pdf>>. Acesso em: 20 Jan. 2016.
- Hall, M. A. **Correlation-based feature subset selection for machine learning**. 1998. 178p. Tese (Doutorado em Filosofia) - University of Waikato, Hamilton - New Zealand, 1998. Disponível em: <<http://www.cs.waikato.ac.nz/~mhall/thesis.pdf>>. Acesso em: 20 Jan. 2016.
- Han, J.; Kamber, M.; Pei, J. **Data mining: concepts and techniques**. 3 ed. Whaltan: Elsevier, 2012. 744p. ISBN(978-0-12-381479-1).
- Hardin, P.; Hardin, A. Hyperspectral Remote Sensing of Urban Areas. **Geography Compass**, v. 7, p. 7-21, 2013.
- Herold, M.; Roberts, D.A; Gardner, E.M.; Dennison, P.E. Spectrometry for urban area remote sensing – Development and analysis of a spectral library from 350 to 2400 nm. **Remote Sensing of Environment**, v. 91, p. 304-319, 2004.
- Huete, A. R. A soil-adjusted vegetation index (SAVI). **Remote Sensing of Environment**, v. 25, p. 53-70, 1988.
- Kohavi, R.; John, G. H. Wrappers for feature subset selection. **Artificial Intelligence**, v. 97, p. 273-324, 1997. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S000437029700043X>>. Acesso em: 20 Jan. 2016.
- Quinlan, J. R. **C4.5: Programs for Machine Learning**. San Mateo: Morgan Kaufmann Publishers, 1993, 300p.
- Rouse, J. W.; Haas, R. H.; Schell, J. A.; Deering, D. W.; Harlan, J. C. **Monitoring the vernal advancement of retrogradation of natural vegetation**. Greenbelt: NASA/GSFC, 1974. 371 p. Disponível em: <<http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19750020419.pdf>>. Acesso em: 20 Jan. 2016.
- Tong, X; Xie, H; Weng, Q. Urban Land Cover Classification With Airborne Hyperspectral Data: What Features to Use?. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v.7, n.10, p.3998-4009, 2014. DOI 10.1109/JSTARS.2013.2272212 Disponível em: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6566095&isnumber=6994884>> Acesso em: 20 jan. 2016.
- Tso, B.; Mather, P. **Classification methods for remotely sensed data**. 2 ed. Boca Raton: Taylor & Francis, 2009. 356 p. ISBN (978-1-4200-9072-7).
- Wang, Y. Y.; LI, J. Feature-selection ability of the decision-tree algorithm and the impact of feature-selection/extraction on decision-tree results based on hyperspectral data. **International Journal of Remote Sensing**, v. 29, n. 10, p. 2993-3010, 2008.
- Walton, J. T. Subpixel urban land cover estimation: comparing cubist, random forests, and support vector regression. **Photogrammetric Engineering and Remote Sensing**, v. 74, n. 10, p. 1213-1222, 2008.