

Avaliação de algoritmos de classificação supervisionada para imagens do Cbers-2 da Região do Parque Estadual do Rio Doce-MG

Mayra Luíza Marques da Silva¹
Geraldo Paulino Marques Pereira²

¹ Universidade Federal de Viçosa - UFV
Av. P.H. Rolfs, s/n. – Campus – 36570-000 - Viçosa - MG, Brasil
mayra_ufv@yahoo.com.br

² Pontifícia Universidade Católica de Minas Gerais – PUC-MG
Av. Dom José Gaspar, 500 – 30535-901 – Belo Horizonte - MG, Brasil
paulino@pucminas.br

Abstract. This paper presents an assessment of algorithms of supervised classification. The classification algorithms utilized were the parallelepiped, minimum distance to means and maximum likelihood. The objective of this work is to evaluate difference it exists among the algorithms had grounds on Kappa Index.

Palavras-chave: remote sensing, algorithms, supervised classification, sensoriamento remoto, algoritmos, classificação supervisionada.

1. Introdução

O sensoriamento remoto, com suas características sinópticas, multiespectrais e temporais, é uma importante ferramenta no mapeamento, identificação e monitoramento de feições terrestres. Possibilitando obter informações sobre a cobertura de áreas extensas, de maneira relativamente rápida, precisa e econômica (Matsukuma, 2002).

As técnicas de processamento digital de imagens permitem tratar os dados e gerar parâmetros adicionais aos métodos convencionais de pesquisa. Dentre estas, a classificação atribui cada pixel da imagem a uma dada classe, a fim de reconhecer padrões e objetos para mapear áreas de interesse. Cada pixel apresenta um valor numérico, ou seja, representa a reflectância dos objetos que o compõem. Como produto final tem-se um mapa temático, o qual representa a distribuição geográfica das classes temáticas.

A classificação apresenta-se de dois tipos: supervisionada e não-supervisionada. A classificação supervisionada depende de amostras de treinamento que sejam representativas das classes presentes na imagem. Assim, o algoritmo classifica os pixels para cada classe. As amostras de treinamento devem ser em número representativas e uniformes. Enquanto que a classificação não supervisionada dispensa a definição do número de classes e das amostras de treinamento, uma vez que o algoritmo agrega pixels e o intérprete identifica as classes geradas pelo algoritmo.

Três dos algoritmos utilizados na classificação supervisionada são: o método paralelepípedo que considera uma área na forma de quadrado ou paralelepípedo ao redor da amostra de treinamento, os lados constituem os chamados limites de decisão; o método da distância mínima até a média que atribui cada pixel à classe cuja média é mais próxima dele; e o método da máxima verossimilhança que utiliza a média e covariância das amostras computando a probabilidade de um pixel desconhecido pertencer a uma ou outra classe.

O desempenho da classificação depende da capacidade de reconhecer as feições e da experiência do interprete com a área de estudo. Porém, qualquer tipo de classificação esta sujeita a erros. Uma vez que o método simplifica a complexidade da imagem, muitas vezes as classes predefinidas são incompletas e apenas as características espectrais são consideradas, deixando de lado as texturais.

A avaliação do desempenho dos algoritmos de classificação digital baseia-se comumente na matriz de erros e do índice Kappa. Neste processo, há a comparação entre a verdade terrestre e os resultados obtidos com a classificação digital. Na matriz de erros podem-se avaliar os pixels corretamente classificados, bem como os erros de comissão e de omissão (Matsukuma, 2002).

O objetivo deste trabalho foi avaliar os diferentes algoritmos de classificação supervisionada a fim de verificar qual fornece a melhor classificação e conseqüente mapa temático. Dessa maneira, obter o mapeamento da área de estudo para usos múltiplos, dentre estes, o monitoramento da cobertura e do uso do solo.

2. Área de estudo

A área de estudo situa-se entre os meridianos 42° 38'W e 48° 28'W e os paralelos 19° 45'S e 19° 30'S, correspondendo ao Parque Estadual do Rio Doce. Com sede na cidade de Marliéria em Minas Gerais e gerenciado pelo IEF (Instituto Estadual de Florestas). A área da unidade de conservação abrange 35.976,43 hectares com um perímetro de 120 km.

O parque compreende os municípios de Timóteo com 14,1% (5.085,26 ha), Dionísio com 2,6% (93.513 ha) e Marliéria com 83,3% (29956,04 ha). Tem como limites o Rio Doce e o Rio Piracicaba. Situa-se no bioma da Mata Atlântica e classifica-se como floresta estacional semidecidual.

Na unidade de conservação são desenvolvidas atividade de educação ambiental, recreação, monitoramento e controle florestal, coleta de sementes e produção de mudas, fomento florestal e pesquisa. E aberto ao público para atividades como banho, caminhada, camping, passeio de barco, trilha, e pesca (IEF).

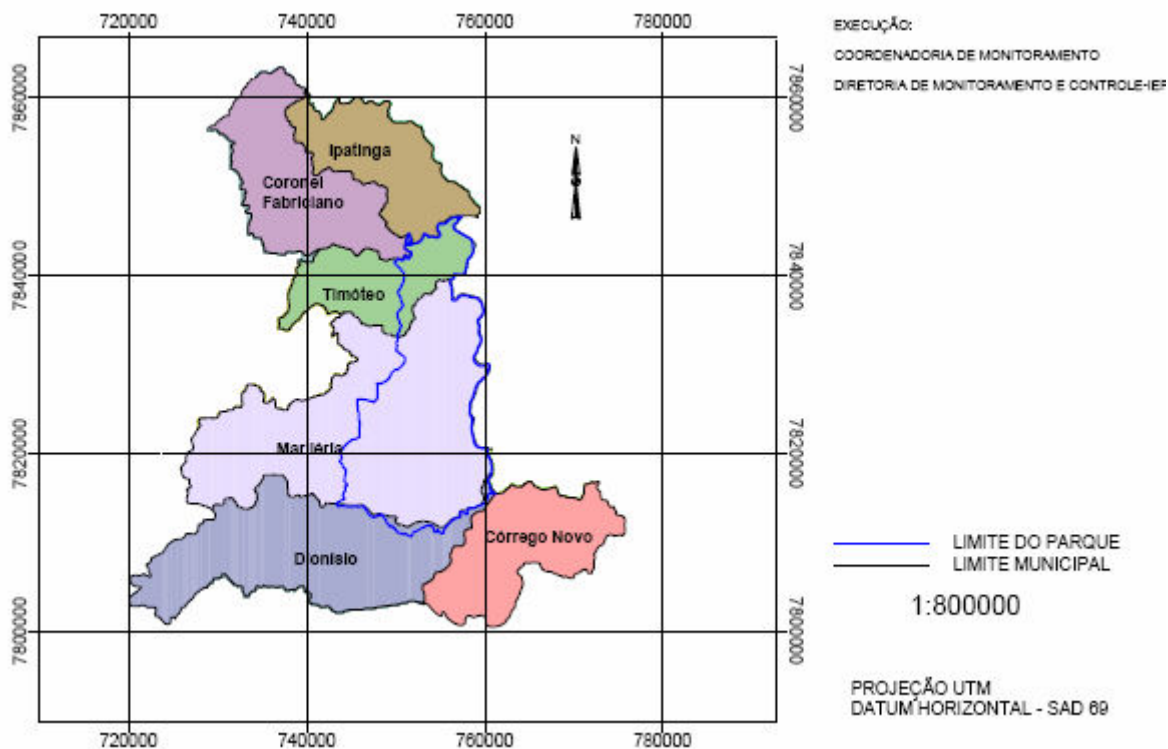


Figura 1 – Parque Estadual do Rio Doce e principais fonte do entorno. (Fonte: IEF)

3. Material e métodos

As imagens utilizadas foram obtidas do sensor CCD do satélite sino-brasileiro, CBERS-2, órbita/ponto 151/122 de 12 de junho de 2006. Foi realizado um corte de 2518 x 1270 pixels, uma vez que o parque localiza-se na porção superior direita da imagem.

As bandas foram georeferenciadas a partir das cartas topográficas 1:100.000 do IBGE dos municípios de Caratinga e Coronel Fabriciano. Posteriormente, foram geradas imagens compostas (2B3G4R) com as bandas CCD 2 (0,52 - 0,59 μm - verde), 3 (0,63 - 0,69 μm - vermelho) e 4 (0,77 - 0,89 μm - infravermelho próximo), a fim de reconhecer as feições terrestres e selecionar as amostras de treinamento, tendo cada classe entre 5 e 10 amostras.

As classes discriminadas na imagem foram: mata (M), reflorestamento (R), pastagem (P), rio (r), lagoa (L) e urbanização (U). Após a coleta das amostras procedeu a extração das estatísticas necessárias para procedimentos posteriores.

O processo de classificação supervisionada envolveu os algoritmos: (a) método do paralelepípedo baseado na média mais ou menos 2 desvios padrões; (b) método do paralelepípedo baseado nos valores máximos e mínimos; (c) método da distância mínima até a média baseado na média normalizada; (d) método da distância mínima até a média baseado na média bruta; e (e) método da máxima verossimilhança. Todo o processamento digital das imagens foi através do software Idrisi32.

A avaliação da imagem classificada baseou-se na coleta de amostras testes numa imagem de referência da região do parque, assumindo serem representativas da verdade terrestre. Os pixels foram confrontados em ambas as imagens gerando a matriz de erros, onde as linhas representam as classes na imagem de referência e as colunas são associadas às classes geradas na classificação. De posse destes dados, obteve-se o índice Kappa, bem como os erros de comissão, ou seja, pixels omitidos de sua correta classe e assinalados em outra classe e erros de comissão, pixels erroneamente incluídos em uma dada classe quando deveria estar em outra (Congalton e Green, 1999).

4. Resultados e discussões

O desempenho dos algoritmos pelo método do paralelepípedo apresentou resultados insatisfatórios, tanto baseado no desvio padrão quanto nos máximos e mínimos. De acordo com a matriz de erros das **Tabelas 1 e 2**, o primeiro algoritmo não classificou 9,58% (1062) dos pixels e o segundo 2,52% (279). E apresentaram índice de Kappa igual a 0,46 e 0,42, respectivamente, conforme **Tabela 6**. Além de altos erros de omissão e comissão.

O método da distância mínima até a média baseada na média normalizada apresentou alto erro de omissão para a classe de mata e de comissão para a classe de reflorestamento devido à semelhança de resposta espectral dos pixels de ambas as classes (**Tabela 3**). Porém, o Kappa obtido teve valor de 0,85 (**Tabela 6**). O mesmo algoritmo baseado na média bruta teve erro significativo na omissão de pixels da classe reflorestamento e kappa 0,95, sendo o melhor valor do índice dentre os algoritmos (**Tabela 6**).

Pelo método da máxima verossimilhança, com erros de omissão elevados para reflorestamento e lagoa e de comissão também para a classe reflorestamento. O índice para avaliação o desempenho do algoritmo, Kappa, foi de 0,89 (**Tabela 6**).

Apesar dos algoritmos de distancia mínima e máxima verossimilhança terem apresentados classificados excelentes de acordo com o Kappa, o método da distância mínima pela média bruta obteve o melhor desempenho. E mesmo as classificações pelo método do paralelepípedo tidas como boas, seus resultados não atendem de maneira precisa para a obtenção de mapas temáticos da região.

O desempenho do algoritmo depende das características da imagem e método de amostragem das amostras de treinamento. A avaliação do algoritmo deve basear-se numa imagem ou mapa de referência com amostras testes que não sobreponham as amostras de treinamento a fim de não superestimar o índice Kappa. Verifica-se que a partir de imagens de satélite e conhecimento de campo, a classificação supervisionada torna-se uma importante ferramenta para o mapeamento, bem como o monitoramento do uso do solo e cobertura terrestre.

Tabela 1 – Matriz de erros para as classes de uso e cobertura do solo geradas pelo método do paralelepípedo baseado na média mais ou menos 2 desvios padrões.

| | M | R | P | r | L | U | Total | Erro C % |
|--------------|-------|------|-------|-------|-------|------|-------|----------|
| Não Classif. | 137 | 45 | 393 | 203 | 179 | 105 | 1062 | 100 |
| M | 541 | 6 | 0 | 0 | 0 | 0 | 547 | 1,1 |
| R | 3923 | 1856 | 0 | 0 | 0 | 0 | 5779 | 67,88 |
| P | 0 | 0 | 1382 | 0 | 0 | 48 | 1430 | 3,36 |
| r | 0 | 0 | 0 | 362 | 0 | 0 | 362 | 0 |
| L | 0 | 0 | 0 | 0 | 436 | 0 | 436 | 0 |
| U | 0 | 0 | 7 | 0 | 0 | 1458 | 1465 | 0,48 |
| Total | 4601 | 1907 | 1782 | 565 | 615 | 1611 | 11081 | |
| Erro O % | 88,24 | 2,67 | 22,45 | 35,93 | 29,11 | 9,5 | | |

Tabela 2 – Matriz de erros para as classes de uso e cobertura do solo geradas pelo método do paralelepípedo baseado nos valores máximos e mínimos.

| | M | R | P | r | L | U | Total | Erro C % |
|--------------|-------|------|-------|-------|-----|------|-------|----------|
| Não Classif. | 0 | 0 | 25 | 221 | 32 | 1 | 279 | 100 |
| M | 7 | 0 | 0 | 0 | 0 | 0 | 7 | 0 |
| R | 4594 | 1904 | 6 | 0 | 0 | 0 | 6504 | 70,73 |
| P | 0 | 3 | 1213 | 0 | 0 | 0 | 1216 | 0,25 |
| r | 0 | 0 | 0 | 340 | 0 | 0 | 340 | 0 |
| L | 0 | 0 | 0 | 0 | 583 | 0 | 583 | 0 |
| U | 0 | 0 | 538 | 4 | 0 | 1610 | 2152 | 25,19 |
| Total | 4601 | 1907 | 1782 | 565 | 615 | 1611 | 11081 | |
| Erro O % | 99,85 | 0,16 | 31,93 | 39,82 | 5,2 | 0,06 | | |

Tabela 3 – Matriz de erros para as classes de uso e cobertura do solo geradas pelo método da distância mínima até a média baseado na média normalizada.

| | M | R | P | r | L | U | Total | Erro C % |
|----------|-------|-------|------|------|-----|------|-------|----------|
| M | 3651 | 192 | 0 | 0 | 0 | 0 | 3843 | 5 |
| R | 950 | 1711 | 0 | 0 | 0 | 0 | 2661 | 35,7 |
| P | 0 | 4 | 1776 | 0 | 0 | 96 | 1876 | 5,33 |
| r | 0 | 0 | 0 | 532 | 0 | 0 | 532 | 0 |
| L | 0 | 0 | 0 | 0 | 615 | 0 | 615 | 0 |
| U | 0 | 0 | 6 | 33 | 0 | 1515 | 1554 | 2,51 |
| Total | 4601 | 1907 | 1782 | 565 | 615 | 1611 | 11081 | |
| Erro O % | 20,65 | 10,28 | 0,34 | 5,84 | 0 | 5,96 | | |

Tabela 4 – Matriz erros para as classes de uso e cobertura do solo geradas pelo método da distância mínima até a média baseado na média bruta.

| | M | R | P | r | L | U | Total | Erro C % |
|----------|------|------|------|-----|-----|------|-------|----------|
| M | 4475 | 206 | 0 | 0 | 0 | 4 | 4685 | 4,48 |
| R | 115 | 1701 | 0 | 0 | 0 | 0 | 1816 | 6,33 |
| P | 0 | 0 | 1767 | 0 | 0 | 50 | 1817 | 2,75 |
| r | 0 | 0 | 0 | 565 | 0 | 3 | 568 | 0,53 |
| L | 11 | 0 | 0 | 0 | 615 | 0 | 626 | 1,76 |
| U | 0 | 0 | 15 | 0 | 0 | 1554 | 1569 | 0,96 |
| Total | 4601 | 1907 | 1782 | 565 | 615 | 1611 | 11081 | |
| Erro O % | 2,74 | 10,8 | 0,84 | 0 | 0 | 3,54 | | |

Tabela 5 – Matriz erros para as classes de uso e cobertura do solo geradas pelo método da máxima verossimilhança.

| | M | R | P | r | L | U | Total | Erro C % |
|----------|------|-------|------|------|-------|------|-------|----------|
| M | 4222 | 203 | 0 | 0 | 0 | 0 | 4425 | 4,59 |
| R | 379 | 1704 | 0 | 0 | 87 | 0 | 2170 | 21,47 |
| P | 0 | 0 | 1771 | 0 | 0 | 107 | 1878 | 5,7 |
| r | 0 | 0 | 0 | 540 | 0 | 0 | 540 | 0 |
| L | 0 | 0 | 0 | 0 | 443 | 0 | 443 | 0 |
| U | 0 | 0 | 11 | 25 | 85 | 1504 | 1625 | 7,45 |
| Total | 4601 | 1907 | 1782 | 565 | 615 | 1611 | 11081 | |
| Erro O % | 8,24 | 10,64 | 0,62 | 4,42 | 27,97 | 6,64 | | |

Tabela 6 – Índice Kappa obtido para os diferentes algoritmos de classificação.

| Algoritmo | Kappa | Qualidade da classificação |
|--------------------------------|-------|----------------------------|
| Paralelepípedo (desvpad) | 0,46 | boa |
| Paralelepípedo (max_min) | 0,42 | boa |
| Distância mínima (norm) | 0,85 | excelente |
| Distância mínima (média bruta) | 0,95 | excelente |
| Máxima verossimilhança | 0,89 | excelente |

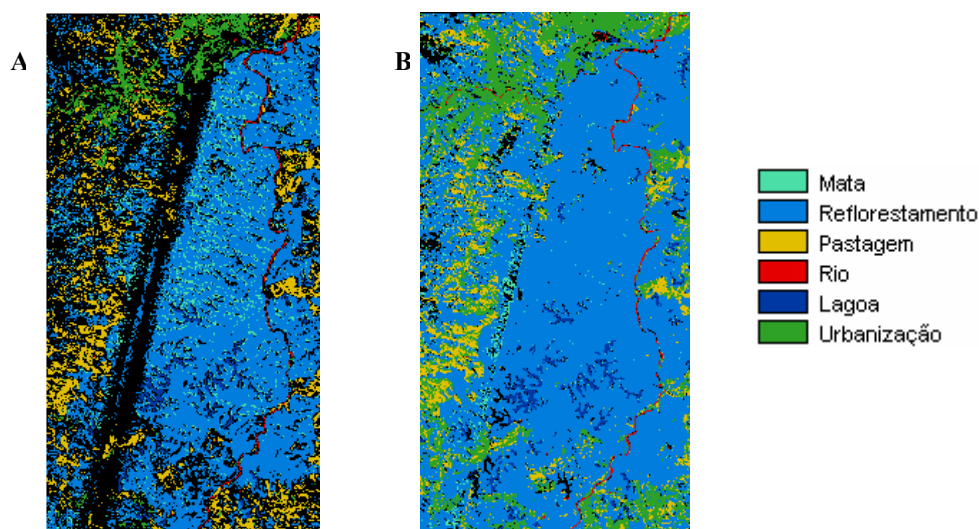


Figura 2 – Imagens classificadas pelo algoritmo do paralelepípedo: a) baseado na média mais ou menos 2 desvios padrões; b) baseado nos valores máximos e mínimos.

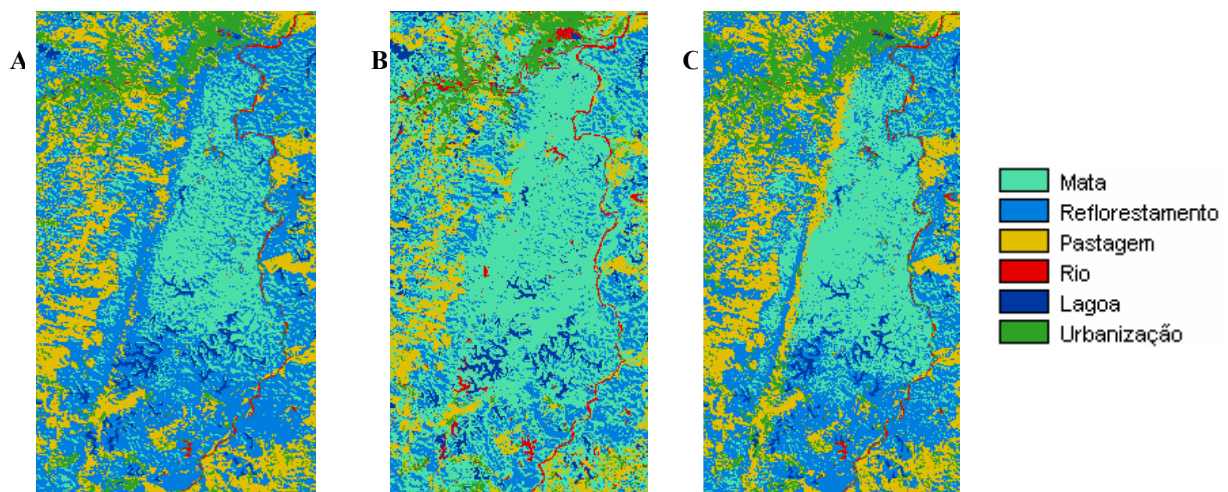


Figura 3 – Imagens classificadas pelo algoritmo da distância mínima até a média: a) Baseado na média normalizada; b) Baseado na média bruta e c) Imagem classificada pelo algoritmo da máxima verossimilhança.

Referências

- Brites, R.S. **Verificação da exatidão em classificação de imagens digitais orbitais: efeitos de diferentes estratégias de amostragens e avaliação de índices de exatidão**. 1996. 101 p. Dissertação (Doutorado em Sensoriamento Remoto) - Universidade Federal de Viçosa. 1996.
- Congalton, R. G. and Green, K. **Assessing the Accuracy of Remotely Sensed Data: principles and practices**. Boca Raton, FL.: Lewis Publishers. 1999. 137 p.
- Gibson, P. J.; Power, C. H. **Introductory Remote Sensing – Digital image processing and applications**. Taylor & Francis, 2000.
- Instituto Estadual de Florestas – IEF. [online] <<http://www.ief.mg.gov.br>>. Outubro. 2006.
- Instituto Nacional de Pesquisas Espaciais - Divisão de Geração de Imagens do (INPE -DGI). Catálogo de Imagens CBERS. Disponível em: <<http://www.dgi.inpe.br/CDSR/>>. Acesso em agosto. 2006.
- Lillesand, T. M.; Kiefer, R.W. **Remote Sensing and Image Interpretation**. United States of America, John Wiley and Sons, 750p., 1994.
- Matsukuma, C. K. **Análise comparativa de algoritmos de classificação não-supervisionada, no mapeamento do uso e cobertura do solo**. 2002. 84 p. Dissertação (Mestrado) – Escola Superior de Agricultura Luiz de Queiroz, Piracicaba. 2002.
- Trotter, C. M. Remotely-sensed data as an information source for geographical information systems in natural resource management: a review. **International Journal Geographical Information Systems**. 5:225-239. 1991.
- Vilela, M. F. **Uso de diferentes métodos de retificação geométrica e classificação digital de uma imagem TM/Landsat-5**. 1998. 118 p. Dissertação (Mestrado em Sensoriamento Remoto) - Universidade Federal de Viçosa. 1998.