



XXXI
CONGRESSO
BRASILEIRO
DE CIÊNCIA
DO SOLO

CONQUISTAS
& DESAFIOS
da Ciência do
Solo brasileira



De 05 a 10 de agosto de 2007 Serrano Centro de Convenções - Gramado-RS

Classificação da cobertura e uso do solo por Redes Neurais Artificiais com o uso de atributos de terreno e imagem de satélite

Eliana de Souza⁽¹⁾; Elídio Inácio Fernandes Filho⁽²⁾; César da Silva Chagas,⁽³⁾ Carlos Ernesto Gonçalves Reynaud Schaefer⁽⁴⁾

RESUMO - A classificação supervisionada por Redes Neurais Artificiais tem apresentado excelentes resultados na classificação de padrões em relação do MAXVER que é um algoritmo largamente utilizado para este fim. O objetivo deste trabalho foi avaliar a classificação supervisionada por RNA das Unidades de Conservação Parque Nacional da Serra do Cipó e Área de Proteção Ambiental Morro da Pedreira, Localizados na Serra do Cipó, MG. De forma específica avaliou a contribuição de atributos do terreno no auxílio aos dados de imagem de satélite no processo de classificação e comparar o desempenho da classificação feita pela RNA e pelo classificador de Máxima Verossimilhança (MAXVER). A classificação foi realizada com dois conjuntos de variáveis. O conjunto 1 (bandas espectrais da imagem Landsat, geologia, MDE – Modelo Digital de Elevação, declividade, ITC – Índice Topográfico Combinado e NDVI - Índice de Vegetação por Diferença Normalizada) e Conjunto 2 (dados da imagem Landsat). Pontos de campo foram adquiridos para avaliar o processo de classificação. Os resultados mostraram que os atributos de terreno aumentam a exatidão da RNA para discriminar as classes de cobertura e uso do solo, com índice kappa de 0,95 para classificação com as variáveis do Conjunto 1 enquanto que só com dos dados da imagem obteve kappa de 0,61. A classificação pelo MAXVER com as variáveis do Conjunto 2 apresentou índice kappa de 0,76, estatisticamente pior que a classificação pela RNA. Concluiu que a abordagem de classificação por RNA é uma técnica de grande potencial para a o mapeamento da cobertura e uso do solo, apresentando melhor desempenho que o MAXVER e que os dados do meio físico são relevantes fontes de informação que auxiliam aos dados de imagem de satélite no processo de classificação.

Palavra chave: Classificação do Uso e cobertura do solo, Redes Neurais Artificiais, Atributos do terreno.

Introdução

Em estudos para a delimitação de Áreas Prioritárias para a conservação em todo o mundo foram definidas 25 áreas denominadas hotspots. No Brasil o Bioma Mata Atlântica, e juntamente com o Bioma Cerrado foram eleitos inclusos nessa categoria, devido à riqueza biológica e à alta pressão antrópica a que vem sendo submetidos. Como mecanismo de conservação da biodiversidade dessas áreas, uma das estratégias tem sido a criação de Unidades de Conservação (UC's) de proteção integral com uso indireto dos recursos ou unidades de uso sustentável com uso direto. UCs O Plano de Manejo (PM) das UC's tem como objetivo orientar o desenvolvimento das atividades nelas desenvolvidas assegurando a manutenção dos recursos naturais. O mapa de uso e cobertura do solo fornece informações de alta relevância para a elaboração do PM, principalmente quando elaborado por processo que garanta alta precisão de mapeamento.

A elaboração do mapa de cobertura e uso do solo consiste no processo de rotular cada pixel da imagem para uma classe informacional. Esse processo é denominado classificação automática quando realizada por algoritmos, ou classificadores (MOREIRA [3]).

O algoritmo de Máxima Verossimilhança (MAXVER) largamente utilizado para classificação supervisionada da cobertura e uso do solo tem suas limitações relacionadas particularmente quanto à exigência da normalidade dos dados (KAVZOGLU & MATHER [1]). Uma abordagem de classificação que procurou superar o problema da normalidade dos dados é fornecida pelas Redes Neurais Artificiais (RNA). Esta abordagem de modelagem utiliza técnicas matemáticas inspiradas no funcionamento de aprendizagem do cérebro humano. O

⁽¹⁾ Mestranda do Departamento de Solos, Universidade Federal de Viçosa. Av. P.H Rolfs, s/n, Departamento de Solos, PPG. Viçosa, MG. CEP 91501-970. elianadsouza@yahoo.com.br.

⁽²⁾ Professor do Departamento de Solos, Universidade Federal de Viçosa. Av. P.H Rolfs, s/n. Sala 123, Campus Universitário. Viçosa, MG. CEP 36570-000. elidio@ufv.br.

⁽³⁾ Pesquisador da Empresa Brasileira de Pesquisa Agropecuária, Centro Nacional de Pesquisa de Solos. Rua Jardim Botânico, 1024. Jardim Botânico. CEP 22460-000 - Rio de Janeiro, RJ. chagas.rj@gmail.com.

⁽⁴⁾ Professor do Departamento de Solos e Nutrição de Plantas, Universidade Federal de Viçosa. Av. P.H. Rolfs s/n. Campus Universitário. Viçosa, MG. carlos.shaefer@ufv.br.

modelo de RNA conhecido como multicamadas e o mais comum, largamente usado, com o algoritmo de aprendizagem *backpropation* (MATHER, [2]).

Considerando a importância dos dados da cobertura e uso do solo para o manejo e conservação este trabalho teve como objetivo avaliar a classificação supervisionada por redes neurais das classes de cobertura e uso do solo, a partir dos dados obtidos pelo sensor ETM+ e de dados do meio físico, com vistas a subsidiar o plano de manejo das. Como objetivos específicos buscou avaliar a contribuição de atributos do terreno no processo de classificação, bem como comparar a os resultados da classificação RNA com o resultado obtido pelo classificador MAXVER.

Material e métodos

A área de estudo compreende as UC's Parque Nacional da Serra do Cipó (PNSCi) e a Área de Proteção Ambiental Morro da Pedreira, situadas na Serra do Cipó, borda oeste da Cordilheira do Espinhaço, Minas Gerais. Essas UC's estão inseridas no Bioma Mata Cerrado, com a presença de extensas formações vegetais do Bioma Mata Atlântica na forma de enclaves de Matas Semidecíduas.

Foram utilizados os seguintes dados: mapa geológico Carta Baldim e Carta Conceição do Mato Dentro (COMIG) Escala 1: 100.000; Mapa Geológico (CPRM). Escala 1:1.000.000; Imagem de satélite da plataforma Landsat 7, órbita/ponto 218-73, adquirida em 23 de abril de 2000, da qual extraiu as bandas 1, 2, 3, 4, 5, e 7 imageadas pelo sensor ETM+ em 23 de abril de 2000; Carta topográfica do IBGE: carta Jaboticatubas, Escala 1:50.000 e cartas Baldim, Conceição do Mato Dentro e Itabira. Escala 1:100.000 e Limites territoriais da área de estudo: PARNA da Serra do Cipó e APA Morro da Pedreira foram fornecidos pelo IBAMA.

Para avaliar a contribuição de atributos de terreno, como variáveis discriminantes no processo de classificação das classes de cobertura e uso do solo, utilizou dois conjuntos de variáveis. O conjunto 1 foi formado por onze variáveis, quais sejam: bandas espectrais da imagem Landsat, geologia, MDE – Modelo Digital de Elevação, declividade, ITC – Índice Topográfico Combinado e NDVI - Índice de Vegetação por Diferença Normalizada. O Conjunto 2 foi composto apenas pelos dados da imagem Landsat.

Utilizou aparelho receptor de Sistema de Posicionamento Global, para a aquisição de pontos em campo e para processamento de dados, classificação dos padrões e apresentação do mapa de cobertura e uso do solo fez-se uso dos seguintes softwares e aplicativos: ERDAS IMAGINE 8.5; *Java Neural Network Simulator* (JavaNNS); Executáveis *FuncPow*, *Transforme* e *Gerapat*; *ArcGIS* 9.2.

As metodologias utilizadas neste trabalho constaram das seguintes etapas: trabalhos de laboratório com implementação de banco de dados em Sistema de Informação Geográfica (SIG), trabalho de campo para reconhecimento de padrões e coleta de amostras das classes

de cobertura e uso do solo, processamento de dados e classificação supervisionada; e avaliação da classificação por meio da matriz de confusão, conforme sugerido por Congalton e Green [3].

A. Trabalho de laboratório

Os trabalhos de processamento dos dados e classificação consistiu na implementação de banco de dados geográfico em Sistema de Informação Geográfica com correção e edição dos dados para corrigir falhas, reduzir a dimensionalidade e realizar transformações para o formato dos softwares utilizados.

B. Trabalho de campo

As incursões à campo realizadas em quatro etapas totalizando vinte dias. Transectos pré-definidos foram percorridos procurando acessar as diferentes tipologias da cobertura vegetal e uso do solo. Foram visitados os diferentes compartimentos da paisagem, com diferentes formas de relevo e unidades litológicas. foram adquiridos 300 pontos georreferenciados por meio de GPS das distintas tipologias de cobertura e uso do solo.

C. Classificação

A arquitetura interna da rede para classificar os três conjuntos de variáveis foi escolhida de forma experimental, tanto o número de camadas quanto o número de neurônios. Para cada conjunto, o número de neurônios na camada de entrada correspondeu ao número de variáveis discriminantes de cada conjunto e o número de neurônios na camada de saída, correspondeu às treze classes a serem classificadas.

D. Definição das classes de cobertura e uso do solo

Após um reconhecimento geral da área, em função da escala e da resolução dos dados disponíveis para o processo de classificação foram identificadas onze classes de cobertura e uso do solo (Tabela 1). Duas outras classes, nuvem e sobra de nuvem, compuseram o arquivo de classificação, por estarem presente na cena da imagem utilizada, totalizando 13 classes informacionais.

Resultados e Discussão

Foi realizado o treinamento das amostras e subsequente a classificação dos pixels dos conjuntos 1 e 2, pela Rede Neural. A avaliação das classificações foi feita por meio de comparação entre as estatísticas das matrizes de confusão. O conjunto de variáveis com maior exatidão na classificação teve as mesmas amostras classificadas pelo algoritmo MAXVER para uma comparação entre classificadores, utilizando as mesmas variáveis, com as mesmas amostras de treinamento e validação.

A. Classificação do conjunto 1 pela RNA

A matriz de confusão obtida para a classificação das variáveis do Conjunto 1 é apresentada na Tabela 1,

com índice kappa foi de 0,95. Observa-se nessa matriz que todas as classes mapeadas apresentaram índice de exatidão acima de 80%. Onze das treze classes apresentaram índice superior a 90% e as demais índice entre 80 e 90%. Os piores desempenhos foram obtidos pelas classes Campo Rupestre herbáceo/arbustivo e Floresta de Galeria com exatidão de 87,5 e 89,0%, respectivamente. Por outro lado, o melhor desempenho, foi obtido pela classe água com índice de exatidão de 99,5%.

B. Classificação do conjunto 2 pela RNA

A matriz de confusão da classificação do Conjunto 2, apresentado na Tabela 2, tem índice kappa de 0,61. Essa classificação apresentou os maiores conflitos entre as classes de Campo Rupestre herbáceo e Campo Rupestre herbáceo/arbustivo. Das treze classes apenas três apresentaram kappa maior que 0,90. Cinco das 13 classes ficaram com um índice entre 0,60 e 0,71 e as demais cinco classes ficaram com o kappa menor que 0,6. As classes com maior exatidão de classificação foram: água, nuvem e sombra com exatidão de 0,99; 0,95 e 0,96 respectivamente. A supressão dos atributos do terreno, neste conjunto acarretou na diminuição fonte de dados. Como consequência houve a diminuição da exatidão de classificação da maioria das classes de cobertura vegetal que ocorrem condicionadas á atributos do terreno. As classes com maior exatidão de classificação possui ocorrência espacial pouco ligada às variáveis do meio físico, sendo, portanto a discriminadas facilmente com a informação espectral da imagem, e com pouca dependência dos dados do meio físico.

C. Classificação do Conjunto 1 pelo MAXVER

A classificação do Conjunto 1 pelo MAXVER obteve índice kappa de 0,74 (Tabela 3). O índice de exatidão por classe mostrou dois extremos de desempenho para quatro das treze classes informacionais. A classe água e a classe Mata Seca não tiveram nenhum pixel classificado corretamente, enquanto, de modo oposto as classes nuvem e sombra tiveram todos os pixels classificados corretamente. O menor desempenho do MAXVER em relação a RNA para o mesmo conjunto de variáveis é explicado método estatístico de tratar os dados. Às variáveis foi aplicado o pelo Teste de *Lilliefors* e observado que de modo geral estes não seguem uma distribuição normal ao nível de probabilidade de 5%.

D. Avaliação do índice de significância entre as classificações

Para observar o nível de significância de cada classificação e comparar o desempenho dos classificadores, utilizou o Teste Z. Este teste determina se dois ou mais valores independentes de kappa e consequentemente duas ou mais matriz de confusão são significativamente diferentes. Os resultados de significância mostraram que as três classificações são estatisticamente diferentes, sendo a

classificação feita pela RNA para o Conjunto 1 melhor do que a classificação para o Conjunto 2, portanto indicando uma melhor precisão da classificação quando usou os dados esta foi feita com os dados da imagem e do meio físico. E a classificação feita pela RNA para o Conjunto 1 melhor do que a classificação feita para este mesmo conjunto pelo MAXVER.

Considerando o resultado das classificações foram concluídas todas as etapas da classificação para o Conjunto 1, com a rotulação de todos os pixels da imagem e filtragem da mesma, para apresentação do mapa (Fig. 1) de cobertura e uso do solo das UC's PNSCi e APA Morro da Pedreira.

Conclusão

Com base nos resultados apresentados conclui-se que a classificação por redes neurais mostrou-se uma técnica adequada para classificar a cobertura e uso do solo das UC's PNSCi e APA Morro da Pedreira, com índice de exatidão geral de 95% para o conjunto que agregou variáveis da imagem de satélite e do meio físico.

Os atributos do terreno são fontes de relevantes informações que auxiliam os dados espectrais de imagem de satélite no processo de classificação. Observou-se a relevância de tais dados na comparação entre a classificação com apenas as bandas espectrais da imagem com kappa de 0,61 (Conjunto 1) e a classificação usando informações das bandas espectrais e atributos do terreno com índice e 0,95 (Conjunto 2).

A classificação pela rede neural mostrou-se mais adequada, do que pelo algoritmo MAXVER, como observados na comparação entre os resultados das matrizes de confusão das mesmas variáveis classificadas pela RNA com exatidão geral de 95% e pelo MAXVER com exatidão geral de 77%. Estes dados reafirmam o melhor desempenho da rede neural em relação ao algoritmo MAXVER, na classificação de padrões que não apresentam distribuição normal.

A classificação da cobertura e uso do solo por rede neural apresenta-se como uma alternativa viável para a elaboração de mapas de cobertura e uso do solo, com o tratamento não paramétrico dos dados permitindo a obtenção de maior exatidão da classificação com o uso de atributos do terreno como auxiliares aos dados de imagem de satélite.

Referências

- [1] KAVZOGLU, T.; MATHER, P.M.. *The use of backpropagating artificial neural network in land cover classification*. International Journal of Remote Sensing. 2003. v.24, p.4907 – 4938.
- [2] MOREIRA, M. [2] A. *Fundamentos do Sensoriamento Remoto: metodologias de Aplicação*. 3ª ed. Viçosa: UFV, 2005. 320p.
- [3] MATHER P. M. *Computer processing of remotely-sensed images: an introduction*. Chichester: John Wiley & Sons. 3ª Ed. 2004. 324 p.
- [4] CONGALTON, R. G. & GREEN, K. *Assessing the accuracy of remotely sensed data: principles and practices*. New York: Lewis Publishers, 1999. 137p.

Tabela 1 – Matriz de confusão do Conjunto 1 classificado pela RNA.

Classes	1	2	3	4	5	6	7	8	9	10	11	12	13	total	Exatidão do analista
1	199	4	0	2	0	0	0	0	0	0	3	0	0	208	95.7
2	1	196	3	0	0	1	0	0	0	1	0	0	6	208	94.2
3	0	0	183	23	3	0	0	0	0	0	2	0	0	214	85.5
4	0	0	12	175	1	0	0	0	0	1	0	0	0	189	92.6
5	0	0	0	0	184	4	1	0	0	0	0	0	0	189	97.4
6	0	0	2	0	0	192	3	2	0	1	6	0	0	206	93.2
7	0	0	0	0	0	0	178	0	0	0	0	0	0	178	100.0
8	0	0	0	0	0	0	0	198	3	0	0	0	0	201	98.5
9	0	0	0	0	7	0	0	0	197	1	0	0	0	205	96.1
10	0	0	0	0	0	0	1	0	0	196	6	0	0	203	96.6
11	0	0	0	0	5	3	1	0	0	0	183	3	1	196	93.4
12	0	0	0	0	0	0	9	0	0	0	0	197	1	207	95.2
13	0	0	0	0	0	0	4	0	0	0	0	0	192	196	98.0
Total	200	200	200	200	200	200	200	200	200	200	200	200	200	2600	
Exatidão do produtor	99.5	98.0	91.5	87.5	92.0	96.0	89.0	99.0	98.5	98.0	91.5	98.5	96.0		

Exatidão global = 95.0; Kappa = 0.946; Variância = 0,000021; Z calculado = 204,278; Z tabelado = 1,96

Classes informacionais: 1-Água; 2-Cerrado – Formações campestre; 3-Campo Rupestre herbáceo; 4-Campo Rupestre herbáceo / arbustivo; 5-Floresta de Candeia; 6-Cerradão - Cerradão + Cerrado sentido restrito; 7-Floresta de Galeria; 8-Floresta Estacional Semidecidual (capão); 9-Floresta Seca; 10-Nuvem; 11-Pastagem; 12-Floresta Estacional Semidecidual; 13-Sombra.

Tabela 2 – Matriz de confusão do Conjunto 2 classificado pela RNA.

Classes	1	2	3	4	5	6	7	8	9	10	11	12	13	total	Exatidão do analista
1	198	3	0	0	0	0	0	0	0	0	0	0	0	201	98,5
2	0	108	40	24	1	2	5	0	0	0	42	3	0	225	48
3	0	26	88	74	2	0	0	0	0	0	4	0	0	194	45,4
4	0	14	35	71	0	0	44	0	0	0	7	0	0	127	55,9
5	0	0	1	0	129	27	17	15	4	0	10	11	0	241	53,9
6	0	0	0	2	36	142	52	3	0	0	1	0	1	199	71,4
7	0	0	3	2	16	17	28	11	8	0	5	0	1	115	45,2
8	0	4	0	0	11	1	41	130	15	0	3	36	0	231	56,3
9	0	0	0	0	5	0	0	33	132	0	0	45	0	256	51,6
10	0	1	0	0	0	0	6	0	0	191	4	0	0	196	97,4
11	0	43	33	27	0	11	7	8	0	6	124	1	0	251	49,4
12	0	0	0	0	0	0	0	0	41	3	0	104	5	168	61,9
13	2	1	0	0	0	0	0	0	0	0	0	0	193	196	98,5
Total	200	200	200	200	200	200	200	200	200	200	200	200	200	2600	
Exatidão do produtor	99	54	44	35,5	64,5	71	26	65	66	95,5	62	52	96,5		

Exatidão global = 63,9; Kappa = 0,609; Variância = 0,000104; Z calculado = 59,863 ; Z tabelado = 1,96

Tabela 3 – Matriz de confusão do Conjunto 1 classificado pelo MAXVER.

[illegible]

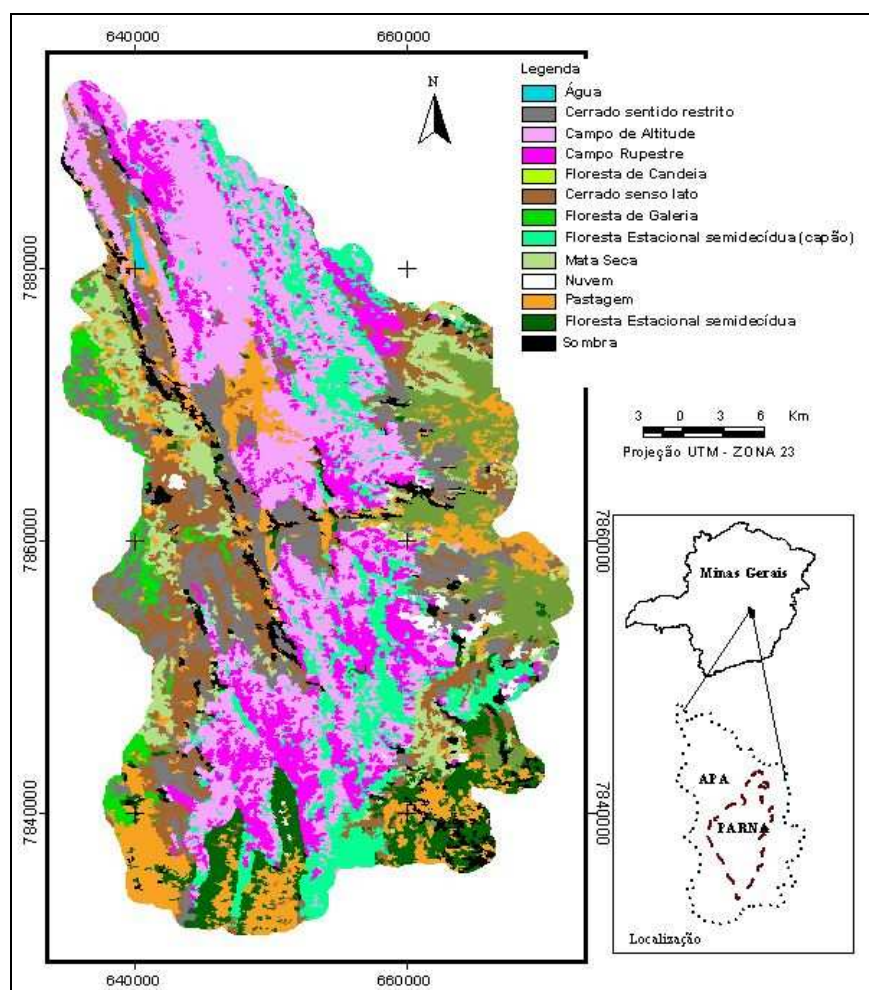


Figura 1 – Mapa de uso e cobertura do solo classificado pela RNA com as variáveis do conjunto 1.