

# COMPARAÇÃO ENTRE CLASSIFICADOR ESTATÍSTICO (MÁXIMA VEROSSIMILHANÇA) E NÃO ESTATÍSTICO (REDES NEURAIIS) VARIANDO O NÚMERO DE CLASSES INFORMACIONAIS

Leonardo Campos de Assis

Giuliano Sant'Anna Marotta

Rômulo Parma Gonçalves

Rafael José de Oliveira Andrade

Mestrandos em Informações Espaciais

Departamento de Eng.<sup>a</sup> Civil, Setor de Eng.<sup>a</sup> Agrimensura, Universidade Federal de Viçosa

Avenida PH Rolfs, Campus Universitário, CEP 36.570-000, Viçosa, MG, Brasil.

{assis\_leonardo, giumarotta, romuloparma, rafaeljoaufv}@yahoo.com.br

Carlos Antônio Oliveira Vieira

Ph.D. em Geografia Física – Sensoriamento Remoto

Departamento de Eng.<sup>a</sup> Civil, Setor de Eng.<sup>a</sup> Agrimensura, Universidade Federal de Viçosa

carlos.vieira@ufv.br

**Resumo** – Dentre as regras de decisão empregadas no sensoriamento remoto, aquelas conhecidas como Classificadoras Hard são bastante utilizadas para grande parte dos propósitos de classificação de imagens obtidas por plataformas orbitais. Entre elas destacam-se a Máxima Verossimilhança (estatístico) e as Redes Neurais Artificiais (não-estatístico). Estes, quase sempre, apresentam melhores resultados quanto à avaliação de desempenho em relação a classificadores Hard convencionais. Por esse motivo, o presente trabalho propôs comparar esses dois classificadores através da avaliação do desempenho dos mesmos alterando o número de classes informacionais e verificando se apresentam diferenças a um nível de confiança de 95% através do teste de Z estatístico. Ao nível de confiança analisado, ambos algoritmos classificadores apresentaram diferenças significativas ao variar o número de classes informacionais. Assim sendo, o critério recomendado para ser adotado quanto à seleção entre um dos classificadores deve considerar o caso em que o mesmo apresentar melhor avaliação de desempenho por meio do coeficiente Kappa. Desse modo as Redes Neurais Artificiais foram recomendáveis em amostras com maior homogeneidade das classes e que apresentaram menor número de classes informacionais, pois, ofereceu melhor resultado quanto ao coeficiente Kappa. Já o algoritmo da Máxima Verossimilhança foi indicado para amostras com maior heterogeneidade, que apresentaram maior número de classes informacionais porque apresentou melhor avaliação quanto ao coeficiente Kappa.

**Palavras-chave** – Regra de decisão, Máxima verossimilhança, Redes neurais artificiais, Classes informacionais.

**Abstract** – About the decision rules employed by remote sensing, that ones knew as Hard Classifiers are largely used by a great number of classification purposes of imagery obtained by orbital platforms. Between them, the Maximum Likelihood (statistical) and the Artificial Neural Networks (non-statistical) are evidenced. These ones present better results on performance evaluation about a great number of common Hard classifiers. For this reason, the present work proposed to compare these classifiers through the performance evaluation of them by differing the number of informational classes and verifying if they present differences at a 95% confidence level through the statistical Z test. At this level, both classifiers algorithms presented significant differences when the number of informational classes differs. In this way, the recommended criterion to be adopted about the selection between one of the classifiers might considerate the case that the classifier presents a better performance evaluation for the Kappa coefficient. Hence, Artificial Neural Networks was recommended on the cases where the samples presented more homogeneity and had a smaller number of informational classes, once it presented a better result for the Kappa coefficient. The Maximum Likelihood algorithm was indicated for samples with more heterogeneity, with a higher number of informational classes, because it presented a better result as evaluation of Kappa coefficient.

**Key-words** – Decision rule, Maximum likelihood, Artificial neural networks, Informational classes.

# 1 INTRODUÇÃO

Classificação de imagens pode ser entendida como o processo de extração de informações de modo a permitir o reconhecimento de padrões e objetos homogêneos com intuito de mapear as áreas da superfície terrestre (Fonseca, 2000). Este processo pode ter duas etapas, a primeira delas relativa ao reconhecimento de categorias de objetos do mundo real e a segunda na identificação de entidades a serem classificadas (Mather, 1999a).

Os tipos de classificação podem ser: supervisionada e não supervisionada. Na classificação supervisionada, faz-se necessário que o usuário conheça algo relativo à área que será classificada antes de iniciar o processo. A classificação não supervisionada fundamenta-se no princípio que o software seja capaz de identificar por si só as classes inseridas num conjunto de dados (Crósta, 1993).

Diversos são os algoritmos disponíveis para classificação supervisionada de imagens, entre os quais merecem destaque: o do paralelepípedo, o da mínima distância e o da máxima verossimilhança.

O algoritmo do paralelepípedo utiliza a meia e o desvio padrão ou os valores máximos e mínimos das amostras de treinamento, já método da mínima distância utiliza a média das amostras de treinamento para assinalar um pixel desconhecido a uma dada classe, atribuindo-o à classe cuja média é a mais próxima (Soares, 2000; Crósta, 1993). O algoritmo da máxima verossimilhança utiliza a média e a covariância das amostras de treinamento, computando a probabilidade estatística de um pixel desconhecido pertencer a uma ou outra classe através de funções de densidade de probabilidade. Após avaliar a probabilidade para cada categoria, o pixel é assinalado à classe cuja probabilidade tenha sido a mais alta, ou mesmo a nenhuma delas para o caso em que os valores de probabilidade estiverem abaixo de certo limiar estipulado pelo usuário (Soares, 2000).

Recentemente, as redes neurais artificiais (*Artificial Neural Network* - ANN) têm sido empregadas como sucesso na classificação de imagens, pois apresentam algumas vantagens sobre os classificadores estatísticos, como por exemplo, o treinamento com um pequeno número de amostras, não ser um método paramétrico, e com grande capacidade de generalização - capacidade de reconhecer com precisão padrões que não participaram do processo de treinamento (Tso e Mather, 2001; Kavzoglu e Mather, 2003).

Independentemente do procedimento adotado, segundo Mather (1999b) há dificuldades inerentes ao princípio de funcionamento dos métodos utilizados nos procedimentos de classificação, como o MaxVer, que admite para as amostras de treinamento uma distribuição normal multivariada. Embora o MaxVer forneça resultados da ordem de 80% ou mais de exatidão, sua confiança em assumir uma distribuição que, por sinal impede o uso de dados categóricos, é motivo para que caia em desuso.

Redes Neurais Artificiais (RNA), alternativamente aos métodos estatísticos, têm sido muito usadas para tipos de dados numéricos, independente do tipo de distribuição desses dados, apresentando ligeiros acréscimos de exatidão em relação aos outros métodos. Entretanto, o desempenho das redes depende fortemente da arquitetura concebida, incluindo o número de camadas internas, o número de neurônios em cada camada interna, a função de transferência associada a cada neurônio na camada interna e o tempo de treinamento da rede por parte usuário através de tentativa e erro (Mather, 1999b).

O objetivo deste trabalho é fazer uma comparação entre os resultados dos procedimentos de classificação dos algoritmos da Máxima Verossimilhança e das Redes Neurais Artificiais através do confronto dos parâmetros de avaliação da exatidão temática (matriz de erros) obtidos para cada um dos métodos pela aplicação do teste estatístico de Z ao nível de confiança de 95%.

## 2 METODOLOGIA

A área de estudo trata-se de uma região situada a nordeste da cidade de Ely, em Cambridgeshire, na Inglaterra. Apresenta características de relevo plano a suave ondulado, com altitude média próxima ao nível dos mares com prática de rotação de culturas agrícolas (Vieira, 2000).

Foram utilizadas três bandas (b1, b2 e b3) de uma imagem SPOT com uma imagem de referência da respectiva área em estudo.

Para realizar uma classificação usando os algoritmos da MaxVer e das RNA; utilizou-se o software Idrisi Kilimanjaro.

A comparação entre os métodos foi realizada aplicando-se o teste estatístico de Z aos valores obtidos do coeficiente Kappa ( $\hat{K}$ ) derivado da matriz de erro produzida pelo cruzamento dos resultados de cada método de classificação com a imagem de referência. Para proceder com a análise comparativa entre os classificadores (regras de decisão), a metodologia de classificação ocorreu em duas etapas, utilizando-se duas imagens de referência, uma delas original com nove classes informacionais e a outra, derivada desta, com três classes (as três mais representativas da área em estudo).

Para submeter as variáveis discriminantes (bandas) ao classificador da máxima verossimilhança, primeiramente, foi necessário realizar a extração de assinaturas ou padrões espectrais das classes categóricas a serem discriminadas. Com o intuito de verificar as semelhanças e/ou diferenças entre as classes utilizou-se um gráfico do tipo *box plot* com as assinaturas espectrais médias de 9 e 3 classes, respectivamente. As classes de categorias utilizadas no procedimento disponíveis na imagem de referência com nove classes são: batatas, beterraba, trigo, fallow, cebolas, ervilhas, bulbos, parsnips e cenouras. A outra imagem de referência apresenta três classes, a saber: batatas, beterrabas e trigo.

Para proceder com a extração das assinaturas, o algoritmo requer a entrada do número mínimo de amostras de pixels por classe, calculado pela equação (1).

$$30 * NDV * NCl \quad (1)$$

onde,  $NVD$  é o número de variáveis discriminantes e  $NCl$  é o número de classes. Portanto, foram selecionados 90 pixels por classe.

Procedeu-se com a discriminação das classes informacionais através da classificação da máxima verossimilhança, empregando a opção de especificar uma probabilidade inicial a priori, para cada assinatura espectral, utilizando a percentagem de omissão na classificação de 0% (classifica todos os pixels). Para se conhecer o valor da probabilidade *a priori* de cada classe calculou-se as áreas (em ha) de todas as classes em ambas as imagens de referência. Tais valores em percentagem foram convertidos para o intervalo de 0 a 1 para serem utilizados como probabilidade.

Quanto ao algoritmo de redes neurais, empregou-se uma mesma arquitetura de rede para ambas as imagens de referência e prosseguiu-se com os testes à procura de um bom desempenho da rede antes da classificação. Dessa forma, os parâmetros de entrada (fator momento, taxa de aprendizado, número de neurônios da camada interna, número de iterações, RMS e taxa de exatidão) para ajuste da rede foram alternados até que se atingiu uma configuração adequada.

A qualidade dos produtos obtidos foi avaliada através da geração da matriz de erros. Através da matriz de erros foi possível determinar, pela equação (2), o valor do coeficiente Kappa, uma vez conhecidos os valores de  $\theta_1$  e  $\theta_2$ , obtidos pelas equações (3) e (4), respectivamente. O coeficiente Kappa trata-se de um índice consagrado na comunidade acadêmica empregado para avaliar o grau de desempenho de uma regra de decisão qualquer

quanto à sua exatidão em relação a uma imagem de referência da mesma área, cujas classes são conhecidas.

$$\hat{K} = \frac{\theta_1 - \theta_2}{1 - \theta_2} \quad (2)$$

onde

$$\theta_1 = \sum_{i=1}^r \frac{x_{ii}}{n} \quad (3)$$

e

$$\theta_2 = \sum_{i=1}^r \frac{x_{i+} x_{+i}}{n^2} \quad (4)$$

nota-se que  $x_{i+}$  é a soma da  $i$ -ésima linha,  $x_{+i}$  é a soma da  $i$ -ésima coluna, e  $x_{ii}$  é a contagem das observações na linha  $i$  e coluna  $i$ . Nota-se também que  $r$  é o número de linhas e colunas na matriz de erros, enquanto o número total de observações é  $n$ .

A variância do coeficiente Kappa ( $\hat{\sigma}_{\hat{K}}$ ) pode ser obtida pela equação (5), desde que conhecidos os valores de  $\theta_3$ , determinado pela equação (6) e  $\theta_4$ , determinado pela equação (7) (Skidmore, 1999).

$$\hat{\sigma}^2(\hat{K}) = \frac{1}{n} \left[ \frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2 - \theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4 - 4\theta_2^2)}{(1-\theta_2)^4} \right] \quad (5)$$

onde

$$\theta_3 = \sum_{i=1}^r \frac{x_{ii}(x_{i+}x_{+i})}{n^2} \quad (6)$$

e

$$\theta_4 = \sum_{i,j=1}^r \frac{x_{ij}(x_{i+}x_{+i})^2}{n^3} \quad (7)$$

Vieira (2000) sugere o teste de Z para determinar diferença significativa entre os valores de coeficientes Kappa independentes, e, portanto, entre matrizes de erro. O teste torna possível a comparação estatística de dois métodos ou algoritmos diferentes. A determinação da distribuição normal do valor de Z é obtida pela razão entre o valor da diferença entre dois coeficientes Kappa ( $\hat{K}$ ) distintos e da diferença dos valores de suas respectivas variâncias ( $\hat{\sigma}_{\hat{K}}$ ). Skidmore (1999) apresenta a equação (8) para obtenção de Z.

$$Z = \frac{\hat{K}_1 - \hat{K}_2}{\hat{\sigma}_{(\hat{K}_1)} - \hat{\sigma}_{(\hat{K}_2)}} \quad (8).$$

### 3 RESULTADOS

Foram comparados graficamente (*box plot*) os valores das assinaturas espectrais médias por classe e por banda, para cada uma das imagens de referência, com nove classes (Figura 1) e com três classes (Figura 2), com o intuito de verificar visualmente possíveis semelhanças e diferenças entre as classes informacionais utilizadas na classificação pelo MaxVer.

Comparando-se as assinaturas espectrais das classes informacionais (Figura 1) pode-se notar que houve maior confusão das assinaturas para todas as variáveis discriminantes ao se considerar a imagem com nove classes. Ao passo que, para o caso da imagem de referência com três classes (

Figura 2), percebe-se uma excelente capacidade de discriminação.

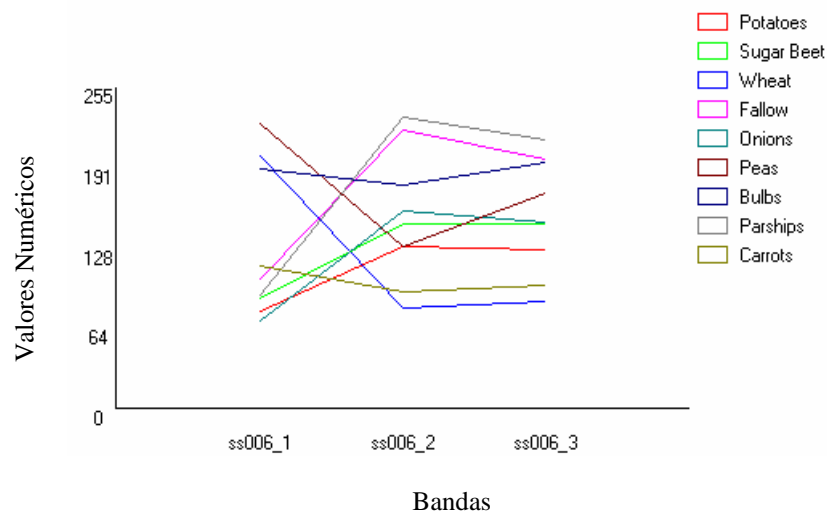


Figura 1: Gráfico de comparação do comportamento de assinaturas das classes categóricas por variáveis discriminantes, considerando-se 9 classes da imagem de referência.

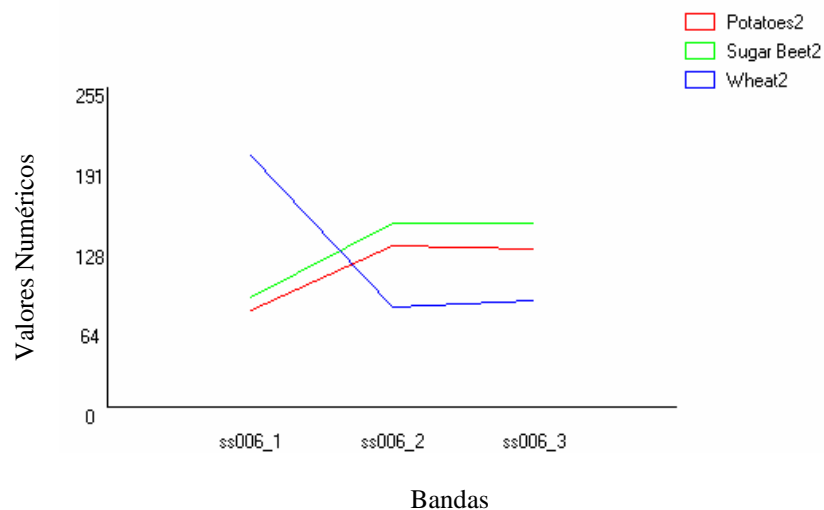


Figura 2: Gráfico de comparação do comportamento de assinaturas das classes categóricas por variáveis discriminantes, considerando-se apenas 3 classes da imagem de referência modificada.

Empregou-se os mesmos parâmetros de configuração para os algoritmos, variando apenas a imagem de referência (com 9 classes e com 3 classes). O algoritmo da MaxVer foi ajustado para classificar todos os pixels (0% de omissão) com definição dos valores de probabilidade *a priori* por classe. Esses valores foram determinados a partir da imagem com nove classes, cujas probabilidades ( $P(x)$ ) são apresentadas na Tabela 1, ao passo que para a imagem com três classes, os valores de ( $P(x)$ ) são apresentados na Tabela 2.

Quanto à melhor configuração encontrada para a rede neural, após sucessivas tentativas de ajuste da arquitetura da rede, a que se comportou melhor apresentou a seguinte configuração: 30 neurônios na camada de entrada, fator momento 0,5 e taxa de aprendizado 0,2.

Tabela 1: Probabilidade de ocorrência  $P(x)$  de uma classe associada à representatividade da mesma na área, calculada para nove classes informacionais da imagem de referência.

Categoria	Área (ha)	%	$P(x)$
Batatas	252.8	14.95823	0.149582
Beterraba	320.92	18.9889	0.189889
Trigo	776.62	45.95276	0.459528
Fallow	84.4	4.993965	0.04994
Cebolas	86.75	5.133015	0.05133
Ervilhas	92.68	5.483894	0.054839
Bulbos	57.48	3.401103	0.034011
Parships	5.81	0.343779	0.003438
Cenouras	12.58	0.744361	0.007444
Total	1690.04	100	1

Tabela 2: Probabilidade de ocorrência  $P(x)$  de uma classe associada à representatividade da mesma na área estudada, para três classes informacionais na imagem de referência.

Categoria	Área (ha)	%	$P(x)$
0	5874.643	81.31013	0.813101
Potatoes	252.8	3.49897	0.03499
Sugar beet	320.92	4.44181	0.044418
Wheat	776.62	10.74909	0.107491
Total	7224.983	100	1

Os valores referentes à avaliação do desempenho dos algoritmos de classificação obtidos pelos coeficientes Kappa e suas respectivas variâncias, referentes ao resultado da classificação submetida às regras de decisão da Máxima Verossimilhança com 0% de omissão na classificação e das Redes Neurais Artificiais com 30 neurônios na camada interna, são apresentados na Tabela 3.

Tabela 3: Valores do coeficiente Kappa ( $\hat{K}$ ) e Variância do coeficiente Kappa ( $\hat{\sigma}_{(\hat{K})}$ ) submetidos à regra de decisão da MaxVer com zero por cento de omissão na classificação e à regra de decisão das RNA com trinta neurônios na camada interna; ambos com nove e três classes informacionais e as mesmas variáveis discriminantes (bandas).

Número de classes	Classificador	$\hat{K}$	$\hat{\sigma}_{(\hat{K})}$
9	MaxVer	0.686	0.000007
	RNA	0.596	0.000007
3	MaxVer	0.722	0.000009
	RNA	0.750	0.000009

A constatação da melhor regra de decisão, para os dados do caso em estudo, foi determinada pela aplicação do teste estatístico de Z (equação 8) ao nível de significância de 95% sobre os valores dos coeficientes Kappa e suas variâncias (Tabela 3) obtidos com cada uma das regras de decisão (MaxVer e RNA), cujos resultados foram os seguintes:

- considerando-se nove classes informacionais, o classificador da MaxVer foi significativamente diferente do classificador das RNA, sendo portanto, recomendado o uso da regra de decisão que apresente o melhor desempenho quanto ao coeficiente Kappa, e, no presente estudo de caso, o melhor desempenho foi conseguido pela regra MaxVer.
- ao se considerar três classes informacionais, o classificador das RNA foi diferente significativamente do classificador da MaxVer, sendo recomendado o mesmo critério de seleção, de uso da regra que apresente melhor desempenho quanto ao coeficiente Kappa, que neste trabalho apresentou melhor resultado para as RNA.

#### 4 DISCUSSÃO E CONCLUSÕES

Os resultados alcançados apresentam comportamento inversamente proporcional dos classificadores da MaxVer e das RNA em relação à quantidade de classes informacionais utilizadas.

O comportamento das redes neurais ocorreu, muito provavelmente, devido à confusão de pixels das classes menos representativas, presente na imagem de referência com nove classes e ausente na imagem com apenas três classes informacionais. Já o algoritmo da máxima verossimilhança apresentou relativamente pouca variação entre os dois métodos, com valores de coeficiente Kappa próximos, tanto para a imagem com nove quanto para a imagem com seis classes.

Os coeficientes Kappa e suas respectivas variâncias foram submetidos também ao teste de Z estatístico, a um nível de significância de 95%, que verificou diferença significativa entre as duas matrizes de erros independentes (MaxVer e RNA).

Os resultados permitem concluir que ambos os classificadores ofereceram grande eficiência durante o processo de classificação, apresentando, no geral, bons valores para os coeficientes de avaliação da classificação. Foi verificado também um semelhante desempenho entre os classificadores. De maneira geral, recomenda-se utilizar as RNA no lugar da máxima verossimilhança em amostras com maior homogeneidade, que sejam mais representativas. Enquanto que, o algoritmo da máxima verossimilhança é mais indicado para amostras com maior heterogeneidade.

## **5 REFERÊNCIAS BIBLIOGRÁFICAS**

CRÓSTA, A. P. Processamento digital de imagens de sensoriamento remoto, Unicamp, Gráfica Universitária 170p. 1993

FONSECA, L. M. G. Processamento digital de imagens, INPE. 2000.

KAVZOGLU, T.; MATHER, P.M. The use of backpropagation artificial neural networks in land cover classification. *International Journal of Remote Sensing*, v.24, n.13, p. 4907-4938. 2003.

MATHER, P.M. *Computer Processing of Remotely-Sensed Images: An Introduction*, second edition. Chichester: Wiley & Sons. 1999a.

MATHER, P. M. Land cover classification revisited. In: *Advances in Remote Sensing and GIS Analysis*. P. M. Atkinson and N. J. Tate, eds. pp. 7-16. Chichester: Wiley & Sons. 1999b.

SKIDMORE, A. K. Accuracy assessment of spatial information. In: A. Stein et. al. (eds.), *Spatial statistics for remote sensing*. p 197-209. Netherlands: Kluwer Academic Publishers. 1999.

SOARES, V. P. Material de aula de sensoriamento remoto, Notas de aula ENF610. Universidade Federal de Viçosa. 73p. 2000.

VIEIRA, C. A. O. Accuracy of remotely sensing classification of agricultural crops: a comparative study. 327p. Thesis (Doctor of Philosophy) – University of Nottingham, Nottingham. 2000.

TSO, B.; MATHER, P. M. Crop discrimination using multi-temporal SAR imagery. *International Journal of Remote Sensing*, vol. 20, no. 12, 2443-2460, 1999.