CLASSIFICAÇÃO PARAMÉTRICA NÃO-SUPERVISIONADA POR MÁXIMA PROBABILIDADE A POSTERIORI

ANTONIO JOSÉ FERREIRA MACHADO E SILVA

CCRio - Centro Científico Rio - IBM Brasil
Av. Presidente Vargas 824 - 22º andar
20071-001, Rio de Janeiro, RJ, Brasil
amachado@riovmsc.vnet.ibm.com
amachado@riosc

Abstract. In this paper we present a model for image classification. This model is a gaussian parametric non-supervised classifier. The approach is to fit the probability distribution function with the image histogram and the discriminant function is one-dimensional.

1. Introdução

O processo de classificação automática de imagens digitais apresenta alguns problemas, dos quais destacam-se:

- modelos paramétricos supervisionados tem a qualidade de seus resultados influenciada pelo treinamento, que serve para determinação dos parâmetros.
- métodos supervisionados não permitem a determinação das probabilidades a priori.
- métodos não-supervisionados geralmente fazem uso de técnicas de agregamento ("clustering"), onde o processo é iterativo, fazendo com que a imagem seja visitada inúmeras vezes. A decorrência imediata é uma elevação substancial nos custos computacionais.
- em alguns modelos, tais como o de máxima probabilidade a posteriori (MAP) e o de máxima verossimilhança (MaxVer), o número de operações de adição e multiplicação é proporcional ao quadrado do número de bandas (NB).

Existem alguns modelos que usam características de métodos supervisionados e não-supervisionados. O modelo MAP iterativo parte de uma classificação MaxVer com parâmetros não necessariamente precisamente conhecidos, e a cada etapa, atualiza tais parâmetros, já incluindo as probabilidades a priori

Este trabalho procura apresentar um processo alternativo, que evita a fase de treinamento, mas não visita a imagem mais que uma vez. O número de operações de adição e multiplicação cresce linearmente com o número de bandas, e além disso determina as probabilidades a priori de cada classe. Em compensação, as covariâncias não são consideradas, o que corresponderia a se ter uma matriz covariância diagonal. Além disso, as funções discriminantes atuam isoladamente sobre cada banda, sendo necessário um processo posterior de tomada de decisão frente aos casos conflitantes.

O modelo procura ajustar um somatório ponderado de funções de distribuição de probabilidade gaussianas aos histogramas de cada banda. Com isso determinar-se-ia as médias e variâncias de cada classe, separadamente para cada banda empregada.

2. Modelo Matemático

Conforme citado anteriormente, o modelo trabalha com as bandas de forma isolada. Isso corresponde a dizer que todo o tratamento matemático é unidimensional.

Para uma dada banda B_k é determinado o histograma de freqüências relativas. Uma primeira observação indica o provável número de classes (NC). A partir daí, procede-se ao ajuste de um somatório ponderado de gaussianas ($F_x(X)$) à função distribuição de probabilidade ($F_x(X)$) baseada no histograma da banda, onde não são conhecidos os parâmetros dessas gaussianas, nem os seus pesos. Os pesos correspondem às probabilidades a priori de cada classe. As equações a seguir procuram ilustrar o problema:

$$\hat{F}_x(X) = \sum_{i}^{\bar{X}} h(i), \quad onde$$

$$\bar{X} = maior\ inteiro \leq X$$

$$h(i) = freq.\ relat.\ do\ nív.\ de\ cinza\ i$$

$$F_x(X) = \sum_{j=1}^{NC} P_j \cdot \int_{-\infty}^{X} f_x^{(j)}(\alpha)\, d\alpha, \quad onde$$

$$f_x^{(j)}(X) = \frac{1}{\sqrt{2\pi}\,\sigma_j} \exp\{-\frac{(X - m_j)^2}{2\sigma_j^2}\}$$

$$F_x(X) = \sum_{j=1}^{NC} P_j \cdot Q(\frac{m_j - Z}{\sigma_j}), \quad onde$$

$$Q(X) = \int_{X}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{X^2}{2}\}$$

Os parâmetros (m_j, σ_j, P_j) formam o vetor de parâmetros **V**. Este vetor desse ser determinado de forma a minimizar o erro médio quadrático ($\mathcal{E}$). No processo de minimização é necessário considerar a restrição de que o somatório das probabilidades a priori tem que ser 1. Em função do que foi exposto, e considerando-se N (= NC) como sendo o número de classes, tem-se:

$$V = (m_1 \dots m_N\, \sigma_1 \dots \sigma_N\, P_1 \dots P_N\, \lambda)^t$$

$$1 - \sum_{j=1}^{NC} P_j = 0$$

$$\mathcal{E}(X) = F_x(X) - \hat{F}_x(X) + \lambda \cdot (1 - \sum_{j=1}^{NC} P_j)$$

$$\mathcal{E} = \int_{-\infty}^{\infty} \mathcal{E}^2(X)\, dX$$

O processo de minimização é iterativo, mas atua apenas sobre os histogramas de cada banda da imagem, sendo bastante rápido. O método escolhido para minimização é o Método de Newton Generalizado ou de segunda ordem, pois utiliza as derivadas parciais de primeira e segunda ordem. Considerando $\nabla\mathcal{E}$ o vetor derivada parcial de primeira ordem e $\nabla^2\mathcal{E}$ a matriz derivada parcial de segunda ordem, e M (= 3.NC + 1) o número de parâmetros a determinar, tem-se:

$$\nabla\mathcal{E} = (\frac{\partial\mathcal{E}}{\partial V_1}\ \frac{\partial\mathcal{E}}{\partial V_2}\ \dots\ \frac{\partial\mathcal{E}}{\partial V_M})$$

$$
\nabla^2 \mathcal{E} = \begin{vmatrix} \frac{\partial^2 \mathcal{E}}{\partial^2 V_1} & \frac{\partial^2 \mathcal{E}}{\partial V_1 \partial V_2} & \cdots & \frac{\partial^2 \mathcal{E}}{\partial V_1 \partial V_M} \\ \frac{\partial^2 \mathcal{E}}{\partial V_2 \partial V_1} & \frac{\partial^2 \mathcal{E}}{\partial^2 V_2} & \cdots & \frac{\partial^2 \mathcal{E}}{\partial V_2 \partial V_M} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 \mathcal{E}}{\partial V_M \partial V_1} & \frac{\partial^2 \mathcal{E}}{\partial V_M \partial V_2} & \cdots & \frac{\partial^2 \mathcal{E}}{\partial^2 V_M} \end{vmatrix}
$$

$$
V_{i+1} = V_i - \delta \, . \, \nabla^2 \mathcal{E}^{-1} \, . \, \nabla \mathcal{E},
$$

onde δ é uma variável utilizada para facilitar a convergência. O valor inicial de cada parâmetro pode ser estimado a partir de algum conhecimento a priori, ou ser estipulado arbitrariamente.

A partir da determinação dos parâmetros, define-se a seguinte função discriminante d_j:

$$
d_j = \ln \sigma_j - \ln P_j + \frac{(X - m_j)^2}{2 \, . \, \sigma_j^2}
$$

Neste caso, decide-se pela classe W_i sempre que $d_i < d_j$, $\forall \; j \neq i$. Esta função discriminante pode ser implementada com 2.NB operações de adição e de multiplicação.

3. A Metodologia de Classificação

O primeiro passo é a estimativa do número de classes que se deseja para o processo de classificação. Este número pode ser estimado por observação das imagens, dos histogramas, ou por conhecimentos a priori que se tenha sobre a região representada na imagem.

Os valores iniciais de média podem representar uma divisão equilibrada do região realmente ocupada no histograma. As variâncias apresentam um comportamento mais equilibrado. Uma vez determinados os parâmetros, calcula-se o erro ($\mathcal{E}$) por integração numérica. Pode-se então variar o número de classes, aumentando ou diminuindo de 1 a 2, repetindo-se o processo para verificação do comportamento do erro.

Este procedimento é realizado tantas vezes quantas forem as bandas, e o primeiro teste de consistência é observar o valor ótimo de número de classes para as diversas bandas, bem como as probabilidades a priori de cada uma.

O processo de classificação gera uma tupla NB-dimensional para cada pixel, onde cada posição da tupla indica a classe na banda correspondente. Para ilustrar, a lista abaixo apresenta algumas decisões do modelo para uma imagem com 5 bandas.

(1,1,1,1,1) → *decide pela classe* 1
(1,2,1,1,1) → *decide pela classe* 1
(1,2,1,1,4) → *decide pela classe* 1
(1,2,1,1,2) → *indecisão* (1 *ou* 2)
(1,2,1,3,2) → *indecisão* (1 *ou* 2)
(1,2,4,3,2) → *indecisão* (1, 2, 3 *ou* 4)
(1,5,4,3,2) → *indecisão* (1, 2, 3, 4 *ou* 5)

Nos casos onde ocorre indecisão, e somente nestes, a decisão pode ser tomada em função da distância euclidiana às médias das classes.

4. Resultados

O modelo foi testado sobre 2 imagens da região amazônica, uma sobre o Estado de Mato Grosso (MT), com 1000 linhas de 1000 pixels cada, e a outra sobre o estado do Pará (PA), com 768 linhas de 768 pixels cada. Para avaliar o resultado foram considerados "gabaritos" obtidos a partir de interpretação visual sobre uma imagem segmentada automaticamente. Além disso foram geradas imagens sintéticas com classes gaussianas, com vetores média e matrizes covariância obtidas da imagem "gabarito"

Na imagem MT foram caracterizadas três classes: floresta (1, peso 51,9%), não floresta (2, peso 24,5%) e desflorestamento (3, peso 23,6%). Na imagem PA foram encontradas três classes também: floresta (1, peso 61,4%), água (2, peso 20,3%) e desflorestemento (3, peso 18,3%). As tabelas 1 e 2 apresentam os resultados encontrados respectivamente para a imagem MT sintética e MT real. As tabelas 3 e 4 apresentam os resultados para as imagens PA sintética e real.

Um ponto na linha I e coluna J destas tabelas indica a percentagem de pixels que estão na classe I na imagem "gabarito" e na classe J na imagem classificada. Assim sendo, a quarta e última coluna apresenta os pesos de cada classe na imagem "gabarito", enquanto que a quarta e última linha indica os mesmos pesos na imagem classificada. O somatório dos pontos da diagonal principal (I=J) indica a percentagem de pixels que não sofreram alterações na classificação.

50,5	1,1	0,3	51,9
0,9	21,6	2,0	24,5
0,3	2,5	20,8	23,6
51,7	25,2	23,1	100,0

Tabela 1: migração de pixels na imagem MT sintética. Erro = 7,1%

49,4	2,4	0,1	51,9
0,2	22,2	2,1	24,5
0,4	3,0	20,2	23,6
50,0	27,6	22,4	100,0

Tabela 2: migração de pixels na imagem MT real. Erro = 8,2%

59,1	1,0	1,3	61,4
0,4	19,4	0,5	20,3
0,8	0,2	17,3	18,3
60,3	20,6	19,1	100,0

Tabela 3: migração de pixels na imagem PA sintética. Erro = 4,2%

60,2	1,0	0,2	61,4
1,9	18,2	0,2	20,3
1,4	0,3	16,6	18,3
63,5	19,5	17,0	100,0

Tabela 4: migração de pixels na imagem PA real. Erro = 5,0%

A tabela 5 procura comparar os resultados destes modelos, em termos de probabilidade de erro, com outros modelos mais clássicos, tais como: de máxima probabilidade a posteriori (MAP, d_1), de máxima verossimilhança (MaxVer, d_1), matriz covariância diagonal (d_2) e da distância euclidiana (d_4). Nesta tabela o modelo apresentado neste trabalho está referenciado como D.

É interessante frisar que nos dois primeiros o número de operações de adição e multiplicação cresce na razão direta do quadrado do número de bandas, enquanto que nos outros dois, bem como no discutido neste trabalho, este número cresce linearmente. Maiores informações sobre a metodologia de avaliação, ou sobre as imagens empregadas consultar [Machado e Silva, 1993].

	MT	MT	PA	PA
	Real	Sint.	Real	Sint.
dl	7,2	5,9	3,5	1,5
d2	7,6	6,2	3,6	1,7
D	8,2	7,1	5,0	4,2
d3	9,0	7,9	7,0	5,7
d4	9,4	8,7	7,6	6,2

Tabela 5: probabilidade de erro: imagem real e sintética.

5. Considerações Finais

Os resultados apresentados não testam o modelo completamente. Não se conseguiu determinar ainda a região de convergência para o processo de minimização. O sistema apresenta instabilidades, principalmente em relação ao parâmetro variância. Um valor inicial discrepante pode impedir a convergência do modelo.

Os casos de indefinições ainda não foram totalmente explorados, e pretende-se investigar alternativas à solução aqui proposta. Cerca de 70 a 80% da imagem está fora da região de indecisão.

Dentre os modelos que apresentam número de operações de adição e multiplicação variando linearmente com o número de bandas, este foi o que teve melhor desempenho. Isto significa que o modelo é rápido e com resultados de classificação satisfatórios.

A grande vantagem do modelo está na eliminação do processo de treinamento, que se não for bem conduzido, pode levar o desempenho dos modelos MAP e MaxVer para perto do aqui proposto.

O processo de ajuste do histograma para determinação das probabilidades a priori (entre outros) pode ser útil para realização do modelo MAP. Neste caso, a matriz covariância é determinada pelo processo natural de treinamento.

7. Referências

Duda, R.O.; Hart, P.E. Pattern Classification and Scene Analysis. John Wiley and Sons, 1973.

Fiacco, A.V.; McCormick, G.P. Nonlinear Programming: Sequential Unconstraibed Minimization Techniques. John Wiley and Sons, 1968.

Fukunaga, K. Introduction to Statistical Pattern Recognition. Academic Press, 1972.

Machado e Silva, A.J.F. Métodos de Avaliação de Modelos de Classificação de Imagens Digitais. Anais do VII Simpósio Brasileiro de Sensoriamento Remoto. Curitiba, PR, maio de 1993.

Richards, J.A. Remote Sensing Digital Image Analysis: An Introduction. Springer-Verlag, 1986.

Van Trees, H.L. Detection, Estimation, and Modulation Theory. John Wiley and Sons, 1968.

Wozencraft, J.M.; Jacobs, I.M. Principles of Communication Engineering. John Wiley and Sons, 1965.