

1. Publicação nº <i>INPE-4206-PRE/1087</i>	2. Versão	3. Data <i>Julho, 1987</i>	5. Distribuição ☐ Interna ☒ Externa ☐ Restrita
4. Origem <i>DPI</i>	Programa <i>PREPROI</i>		
6. Palavras chaves - selecionadas pelo(s) autor(es) <i>RECONHECIMENTO ESTATÍSTICO DE PADRÕES</i> <i>AGREGAÇÃO DE DADOS</i> <i>SENSORIAMENTO REMOTO</i> <i>EXTRAÇÃO DE ATRIBUTOS</i>			
7. C.D.U.: <i>621.376.5</i>			
8. Título <i>BREVE INTRODUÇÃO AO RECONHECIMENTO ESTATÍSTICO DE PADRÕES</i>		10. Páginas: <i>12</i>	
		11. Última página: <i>11</i>	
		12. Revisada por	
9. Autoria <i>Nelson Delfino D'Ávila Mascarenhas</i>		<i>Flávio Roberto D. Velasco</i> Flávio Roberto D. Velasco	
		13. Autorizada por <i>Dr. Marco Antonio Raupp</i> Dr. Marco Antonio Raupp Diretor Geral	
Assinatura responsável <i>Nelson Mascarenhas</i>			
14. Resumo/Notas <i>Neste trabalho é apresentada uma descrição sumária das principais técnicas de reconhecimento estatístico de padrões. Discutem-se tópicos como: extração de atributos, classificação estatística, métodos de aprendizado e agregação de dados. São apresentados também os principais problemas abordados atualmente na aplicação dessas técnicas na área de detecção de recursos naturais.</i>			
15. Observações <i>Trabalho submetido para apresentação na 39ª Reunião Anual da SBPC, Simpósio: "Inteligência Artificial", em Julho/87, Brasília.</i>			

BREVE INTRODUÇÃO AO RECONHECIMENTO ESTATÍSTICO DE PADRÕES

Nelson Delfino d'Ávila Mascarenhas
Ministério da Ciência e Tecnologia - MCT
Instituto de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12201 - São José dos Campos - SP

RESUMO

Neste trabalho é apresentada uma descrição sumária das principais técnicas de reconhecimento estatístico de padrões. Discutem-se tópicos como: extração de atributos, classificação estatística, métodos de aprendizado e agregação de dados. São apresentados também os principais problemas abordados atualmente na aplicação dessas técnicas na área de detecção de recursos naturais.

ABSTRACT

In this work a brief description of the main techniques of statistical pattern recognition is presented. The discussed topics comprise: feature extraction, statistical classification, learning methods and clustering. The main problems tackled presently in the area of natural resources detection are also presented.

BREVE INTRODUÇÃO AO RECONHECIMENTO ESTATÍSTICO DE PADRÕES

A área de reconhecimento de padrões não está restrita à classificação e descrição de imagens, mas a maior parte do trabalho desenvolvido tem-se voltado para aplicações pictóricas. Assim sendo, tem havido tentativas sérias de automatizar as tarefas de leitura de caracteres alfanuméricos, análise de imagens aeroespaciais, biomédicas e de câmaras de bolha em física de alta energia, bem como imagens obtidas por robôs industriais. Destacam-se também aplicações não-pictóricas como reconhecimento de voz e do orador, análise de sinais de sonar, eletrocardiograma, eletroencefalograma e em sismologia, assim como classificação de dados obtidos em antropologia ou zoologia, etc.

O reconhecimento de padrões tem-se desenvolvido, em geral, ao redor de duas grandes linhas de abordagem:

- 1) Classificação de padrões, onde o objetivo é atribuir um objeto a uma das possíveis classes. A ferramenta básica é a teoria da decisão estatística, embora várias técnicas empregadas sejam essencialmente determinísticas.
- 2) Reconhecimento sintático (ou estrutural) de padrões, onde se procura uma descrição do objeto em termos de relações entre suas partes, utilizando a teoria de linguagens formais. Este trabalho vai abordar apenas as técnicas de classificação de padrões que têm sido mais utilizadas em áreas como detecção de recursos naturais.

Os conhecimentos atuais de neurofisiologia não têm sido suficientes para reproduzir, na máquina, as funções de percepção de seres muito primitivos. Todavia, mesmo diante desta insuficiência, tem havido êxito na resolução de diversos problemas de reconhecimento de padrões, particularmente nas tarefas mais simples de classificação, ou seja, a de atribuir um objeto a uma de diversas classes. Os métodos desenvolvidos têm tido por base, em geral, uma formulação estatística que pode estar distante dos processos perceptuais em seres vivos.

Deste modo, é feito, inicialmente, um conjunto de medidas por um sensor. No caso de imagens, estas medidas tendem a se adaptar

ao modo de aquisição das imagens e ao tipo de classificação que se planeja desenvolver. Imagens de satélites de recursos naturais são colhidas em geral sobre várias bandas do espectro (tipicamente de 4 a 12) fornecendo assim um atributo natural para a classificação de um ponto no solo através do vetor de medidas nessas bandas. Outras possíveis medidas consistiriam em medidas de textura ou de tamanho e forma de objetos nessas imagens. Os dados obtidos são representados por meio de um vetor de medidas num espaço com uma determinada dimensão. Normalmente, é necessário efetuar um processo de redução de dimensionalidade (seleção de atributos) antes da fase de classificação, de modo a reduzir o esforço computacional nesta última fase.

A etapa de classificação reduz-se essencialmente a uma partição do espaço de atributos de tal maneira que, idealmente, os pontos pertencentes a uma determinada classe caiam sempre na partição correspondente. Como esse objetivo ideal nem sempre pode ser alcançado, deve-se estabelecer a partição minimizando algum critério que, de acordo com a formulação estatística do problema, é a probabilidade de erro ou um risco médio, dependendo de custos apropriadamente escolhidos. Quando as classes apresentam pequena variação em torno de um protótipo, a partição pode ser efetuada de modo a atribuir o objeto ao protótipo mais próximo. Contudo, em muitas situações, a variação é apreciável e técnicas mais sofisticadas de decisão estatística devem ser empregadas. Essas variações podem ser tanto devidas à incerteza inerente à caracterização da classe, como oriundas de perturbações estatísticas (ruído) nos sensores de medida.

O problema de classificação de padrões está intimamente ligado ao problema de aprendizado estatístico, uma vez que nas situações reais é raro ter uma descrição estatística completa das densidades de probabilidades que caracterizam cada classe.

As técnicas de extração de atributos podem ser consideradas como uma forma de compressão de dados, devido ao caráter de redução de dimensionalidade que elas possuem. Estas técnicas podem ser divididas: a) Transformações no Espaço de Atributos. O objetivo aqui é transformar o espaço de atributos originais em um espaço de menor dimensionalidade, preservando ao máximo a representação dos padrões e/ou a descri

minação entre as classes. No primeiro caso, têm sido usados critérios de erro médio quadrático ou entropia. A Transformação de Karhunen-Loeve, envolvendo os vetores e os valores próprios da matriz de covariância, desempenha um importante papel nesta área. Quanto à discriminação entre classes, é possível utilizar transformações lineares ou não-lineares, sendo obtido um resultado superior no último caso, ao preço todavia de uma carga computacional que pode ser bastante superior. Frequentemente é necessário usar técnicas iterativas ou interativas, através de um "display" (Chien, 1978). Pode haver conflito entre os critérios de representação e os de separabilidade entre classes, e seria desejável que se estabelecesse um compromisso entre esses fatores na escolha da transformação ótima. b) Medidas de Informação e Distância. Em última análise, o critério a ser utilizado para avaliar o desempenho de um sistema de classificação deveria ser a probabilidade de erro, mas existem grandes dificuldades em empregar esse critério diretamente na redução da dimensionalidade. Assim sendo, utilizam-se critérios indiretos expressos por medidas de distância entre distribuições, os quais fornecem limites superiores e/ou inferiores para a probabilidade de erro. Várias medidas têm sido propostas e investigadas as relações entre elas (Kanal, 1974). A divergência e a distância de Bhattacharyya têm a vantagem de ser mais fáceis de computar do que outras medidas propostas.

O problema central de classificação, para um dado vetor de atributos observado, envolve basicamente o cômputo dos valores das funções densidade de probabilidade, relativas às várias classes, e a determinação do máximo valor, a cuja classe é atribuído o referido vetor. Quando as probabilidades a priori são conhecidas e os custos de decisão são escolhidos, o problema enquadra-se na chamada teoria bayesiana, ao passo que se as classes são equiprováveis tem-se a decisão por máxima verossimilhança.

O modelo gaussiano tem sido utilizado frequentemente para caracterizar as classes, particularmente na área de sensoriamento remoto (Crane et alii, 1972). Em geral, nestes casos, um limiar é também estabelecido para rejeitar padrões cuja distância ponderada - a distância de Mahalanobis - a cada uma das classes é superior ao limiar. Também têm sido utilizados os chamados classificadores de paralelepípedos

em que cada classe é definida pelos limites inferior e superior do histograma obtido na área de treinamento. Tais classificadores podem ser implementados de maneira muito rápida por "hardware", mas seu desempenho tem deficiências, principalmente devido à ambiguidade que resulta da sobreposição de paralelepípedos relativos a duas ou mais classes.

A possibilidade de utilizar os esquemas de testes sequenciais de hipóteses tem sido explorada em classificação de padrões, dada a limitação por vezes existente no número de amostras disponíveis. Fu (1968) apresenta vários resultados da aplicação desta técnica na classificação de caracteres alfanuméricos manuscritos.

Os métodos de classificação descritos a seguir referem-se essencialmente ao problema que ocorre quando as densidades de probabilidade que caracterizam as classes não são conhecidas completamente.

Quando a forma funcional das densidades é conhecida, exceto por seus parâmetros, daí resulta um problema de aprendizado que é um processo de estimação dos parâmetros desconhecidos. Duas situações diferentes podem ocorrer, dependendo do tipo de amostras disponíveis: a) amostras de classificação conhecida; b) amostras de classificação não conhecida. No primeiro caso, tem-se o chamado aprendizado com supervisão, e no segundo, o aprendizado sem supervisão.

No aprendizado com supervisão, a disponibilidade de um rótulo nas amostras simplifica consideravelmente o esforço computacional. Se os parâmetros são considerados não-aleatórios, as técnicas de estimação por máxima verossimilhança podem ser empregadas. Já na situação em que os parâmetros são considerados variáveis aleatórias cai-se num esquema bayesiano. Como os estimadores dos parâmetros contêm erros, o desempenho do classificador não é tão bom como no caso em que os parâmetros são conhecidos. Todavia, com a disponibilidade de um grande número de amostras a qualidade dos estimadores melhora, assim como o desempenho do classificador.

Outra situação que ocorre frequentemente é quando a forma das densidades de probabilidades não é conhecida. As formas paramétricas usuais raramente descrevem com fidelidade as densidades encontradas na prática. Em geral, estas formas paramétricas têm a forma unimodal,

enquanto os histogramas reais são muitas vezes multimodais. Deste modo, recorre-se às chamadas técnicas não-paramétricas. Algumas destas técnicas envolvem a estimação das densidades condicionais de cada classe a partir das amostras. Outras técnicas estimam diretamente as probabilidades a posteriori das classes que são usadas como funções discriminantes, sem passar pelas densidades condicionais. Estes procedimentos estão relacionados com o método de classificação de uma amostra pelo protótipo de classificação conhecida mais próximo (vizinho mais próximo). Em geral, as técnicas não-paramétricas trocam a necessidade da forma funcional das densidades ser conhecida pela necessidade de um número maior de amostras.

Suponha-se agora que sejam oferecidas ao classificador apenas amostras de classificação conhecida. Nestas condições o classificador deve ajustar iterativamente as superfícies de decisão de modo a, se possível, separar perfeitamente as amostras. Estes procedimentos, que tiveram muita popularidade no início da década de 60, são os algoritmos do tipo "perceptron" ou de relaxação. Conforme já foi mencionado, tais algoritmos também têm sido utilizados em reconstrução de imagens por projeções, quando se substitui o problema da solução de um sistema linear de igualdades por um sistema linear de desigualdade. Para lidar com a situação em que as classes não são separáveis linearmente, foram desenvolvidos os chamados procedimentos de mínimo erro médio quadrático (Ho and Kashyap, 1965), que convergem ainda na situação não-separável. A função objetivo que caracteriza os algoritmos do tipo "perceptron" também pode ser minimizada por métodos de programação linear. Devem ser mencionadas também, entre as funções discriminantes lineares, as chamadas funções de potencial, cujo desenvolvimento se inspirou nos métodos da Física Clássica.

Se a única informação possível para projetar o classificador consiste em amostras não-rotuladas, recorre-se às técnicas de agregação de dados ("clustering"). Essas técnicas tiveram origem nos processos de taxonomia numérica das Ciências Biológicas, e com o advento do computador digital passaram a ser largamente utilizadas. Em geral, os métodos de agregação são não-supervisionados e até mesmo o número de classes pode não ser determinado a priori. As técnicas desenvolvidas ba

seiam-se frequentemente na minimização de um critério derivado de uma medida de similaridade entre amostras. Este critério pode ser mínima varriância, envolvendo as matrizes de espalhamento entre classes ou dentro de classes. Técnicas baseadas em teoria de grafos, na análise combinatória ou ainda na teoria de conjuntos nebulosos ("fuzzy sets") têm sido desenvolvidas. No caso de imagens, um procedimento bastante utilizado é o chamado algoritmo de K-médias, cujas propriedades de convergência foram estudadas por Mac Queen (1967). O algoritmo denominado Isodata (Ball and Hall, 1965) representa um desenvolvimento do algoritmo de K-médias, incorporando procedimentos heurísticos para subdividir ou agrupar aglomerados. Deve-se observar que tanto o algoritmo de K-médias quanto o Isodata têm sido utilizados para quantização tanto no caso unidimensional (Velasco, 1980) (sendo equivalente ao procedimento de (Max, 1960)) como no caso vetorial (Linde et alii, 1980).

Os procedimentos de agregação podem ser classificados como hierárquicos ou não-hierárquicos. Nos procedimentos hierárquicos, as partições das amostras são feitas em níveis, por aglomeração ou divisão. A característica hierárquica baseia-se na propriedade de que se duas amostras estão num mesmo agregamento num nível do processo elas assim permanecerão até o fim do processo. Os exemplos mais conhecidos de agregação hierárquica ocorrem em Biologia, onde indivíduos se agrupam em espécies, espécies em gêneros, gêneros em família, etc. A despeito de sua enorme importância prática, os métodos de agregação frequentemente se ressentem de uma base mais sólida nas decisões de dividir ou aglomerar amostras, ou na escolha do número de aglomeramentos. Assim sendo, desenvolve-se atualmente um esforço considerável no sentido de validar estatisticamente estes processos.

É usual em problemas de classificação de padrões adotar a hipótese de que na fase de aprendizado dos parâmetros estes estejam estimados com um número ilimitado de amostras. Todavia, isso nem sempre corresponde à realidade. A frequente disponibilidade de apenas um pequeno número de amostras pode afetar bastante o desempenho do classificador. É uma área ativa de pesquisa o estudo do relacionamento que deve existir entre a dimensionalidade das amostras e seu número, e de como isto afeta a tarefa de classificação.

Em última análise, o desempenho do classificador é medido pela probabilidade de erro de classificação. O procedimento de avaliar esta probabilidade pela classificação do próprio conjunto de amostras de treinamento é considerado otimista, de modo que é necessário dividir o conjunto de amostras numa parte de treinamento e numa parte de teste. Quando se dispõe de um pequeno número de amostras, é preferível fazer a estimativa da probabilidade de erro em N passagens, em cada uma das quais uma observação é mantida para teste e as restantes N-1 são usadas para treinamento.

Para completar este breve sumário das aplicações de classificações de padrões, devem ser mencionados alguns desenvolvimentos recentes e tendências futuras, particularmente no caso de imagens de recursos naturais (Landgrebe, 1981; Swain, 1985). As imagens de sensoramento remoto são colhidas em geral em várias bandas do espectro e, como consequência, os métodos tradicionais de classificação são baseados em características espectrais. Tem havido recentemente um grande esforço para desenvolver algoritmos incorporando atributos espaciais, variações temporais e dados auxiliares. Dentre as técnicas incorporando atributos espaciais podem ser mencionadas: a) filtragem espacial para extrair atributos locais acrescidos aos atributos espectrais (Dondes and Rosenfeld, 1982; Dutra and Mascarenhas, 1984); b) análise de textura por meio de modelos de processos aleatórios bidimensionais (Therrien et alii, 1986) ou por matrizes de co-ocorrência, que são essencialmente histogramas bidimensionais dos valores dos "pixels" para uma dada distância e um dado ângulo; c) algoritmos de partição de cenas, classificando "pixels" por grupos, através de processos de anexação de regiões, por critérios estatísticos (Kettig and Landgrebe, 1976); d) teoria de decisão composta que leva em conta a correlação de "pixels" adjacentes na classificação (Swain et alii, 1980; Chittineni, 1981); métodos de relaxação (Kittler and Illingworth, 1985), que são processos iterativos de atualização das probabilidades a posteriori das classes baseados na dependência probabilística com "pixels" vizinhos.

Os algoritmos incorporando características temporais utilizam-se da periodicidade da cobertura de uma dada região pelos satélites de detecção de recursos naturais. Essas técnicas pressupõem o re

gistro de passagens obtidas pelo satélite em datas diferentes. A primeira abordagem utilizada simplesmente adicionava os atributos espectrais de novas passagens aos atributos da primeira passagem. Todavia, surgem problemas devidos à necessidade de um número maior de amostras de treinamento como decorrência da maior dimensionalidade das amostras, bem como aqueles devidos à não-disponibilidade de dados por cobertura de nuvens. Têm sido estudados métodos que envolvem o estudo das trajetórias temporais espectrais de tipos de terreno (Misra and Wheeler, 1978) (particularmente em levantamento de safras) bem como uma abordagem bayesiana num classificador em cascata (Swain, 1978).

A utilização de dados auxiliares pode-se basear em informações subjetiva e descritiva, como práticas de cultivo ou condições de precipitação pluvial na área ou em dados quantitativos, como perfil topográfico, resposta a outros sensores como radar, mapas de classificação de anos prévios, fronteiras políticas, etc. Um exemplo de uso de dados auxiliares na classificação é aquele que foi utilizado no chamado Procedure 1 do LACIE ("Large Area Crop. Inventory Experiment"), um extenso programa de levantamento de safras de trigo desenvolvido por várias agências nos EUA. Nesse caso o analista dispunha, no processo de rotulação de pontos selecionados dos segmentos a ser classificados, de vários dados auxiliares, como plotagem de trajetórias temporais em espaços transformados (Kauth and Thomas, 1976). Deve ser mencionada finalmente uma tendência crescente na análise de recursos naturais, que consiste na utilização dos chamados sistemas geográficos de informação, que incorporam dados auxiliares que, associados em geral as imagens de satélites, fornecem informações para o gerenciamento de recursos naturais.

REFERÊNCIAS BIBLIOGRÁFICAS

- BALL, G.H.; HALL, D.J. "Isodata, an Iterative Method of Multivariate Analysis and Pattern Classification", Proceedings of the IFIPS Congress, 1965.
- CHIEN, Y.T. "Interactive Pattern Recognition", New York, NY, Marcel Dekker, 1978.

- CHITTINENI, C.B. "Utilization of Spectral-Spatial Information in the Classification of Imagery Data", *Comp. Graphics Image Proces.*, vol. 16, nº 4, Agosto 1981, pp. 305-340.
- CRANE, R.B.; MALILA, W.A.; RICHARDSON, W. "Suitability of the Normal Density Assumption Likelihood for Processing Multispectral Scanner Data", *IEEE Trans. Geosci. Electron.*, vol. Ge-10, nº 4, 1972, pp. 158-165.
- DONDES, P.A.; ROSENFELD, A. "Pixel Classification Based on Gray Level and Local 'Busyness'", *IEEE Trans. Pat. Anal. Mach. Intel.* vol. PAMI-4, nº 1, Janeiro 1982, pp. 79-84
- DUTRA, L.V.; MASCARENHAS, N.D.A. "Some Experiments with Spatial Feature Extraction Methods in Multispectral Classification, *International Journal of Remote Sensing*, vol. 5, nº 2, pp. 303-313, 1984.
- FU, K.S. "Sequential-Methods in Pattern Recognition and Machine Learning", New York, NY, Academic, 1968.
- HO, Y.C.; KASHYAP, E.L. "An Algorithm for Linear Inequalities and its Applications", *IEEE Trans. Elec. Comp.* vol. EC-14, Outubro de 1965, pp. 683-688.
- KANAL, L. "Patterns in Pattern Recognition: 1968-1974", *IEEE Trans. Inform. Theory*, vol. IT-20, nº 6, Novembro 1974, pp-697-722.
- KAUTH, R.J.; THOMAS, G.S. "The Tasselled Cap - A Graphic Description of the Spectral - Temporal Development of Agricultural Crops as seen by LANDSAT" in *Proc. 1976 Symp. Machine Processing of Remotely Sensed Data*, West Lafayette, IN, Junho 1976.
- KETTIG, R.L. LANDGREBE, D.A. "Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects", *IEEE Trans. Geosci. Electron.* vol. GE-4, Janeiro 1976.
- KITTLER, J.; ILLINGWORTH, J. "Relaxation Labelling Algorithms - A Review", *Image and Vision Computing*, vol. 3, nº 4, Novembro 1985, pp. 206-216.
- LANDGREBE, D.A. "Analysis Technology for Land Remote Sensing, *Proc. IEEE*, vol. 69, nº 5, Maio 1981, pp. 628-642.
- LINDE, Y.; BUZO, A.; GRAY, R.M. "An Algorithm for Vector Quantizer Design", *IEEE Trans. on Commun.* vol. COM-28, nº 1, Janeiro 1980, pp. 84-95.

- MAC QUEEN, J. "Some Methods for Classification and Analysis of Multivariate Observations", in Proc. Fifth Berkeley Symposium on Math. Stat. and Prob., I, Le Cam, L.M.; Neyman, J. (Eds.), Berkeley and Los Angeles, Calif., Univ. of California Press, 1967, pp. 281-297.
- MAX, J. "Quantizing for Minimum Distortion", IRE Trans. Inf. Theory, vol. IT-6, nº 1, Março 1960, pp. 7-12.
- MISRA, P.N.; WHEELER, S.G. "Crop Classification with LANDSAT Multispectral Scanner Data", Pattern Recognition, vol. 10, Janeiro 1978.
- SWAIN, P.H. "Bayesian Classification in a Time-Varying Environment", IEEE Trans. Syst., Man., Cybern., vol. SMC-8, Dezembro 1978.
- SWAIN, P.H. "Advanced Interpretation Techniques for Earth Data Information Systems", Rec. IEEE, vol. 73, nº 6, Junho 1985, pp. 1031 - 1039.
- SWAIN, P.H.; SIEGEL, H.J.; SMITH, B.W. "Contextual Classification of Multispectral Remote Sensing Data Using a Multiprocessor System", IEEE Trans. Geosci. Remote Sensing, vol. GE-18, Abril 1980.
- THERRIEN, C.W.; QUATIERI, T.F.; DUDGEON, D.E. "Statistical Model-Based Algorithms for Image Analysis" Proc. IEEE, vol.74, nº 4, Abril 1986, pp. 532-551.

BIBLIOGRAFIA

- AGRAWALA, A.K., ed. Machine recognition of patterns. New York, NY, IEEE Press, 1976.
- ANDREWS, H.C. Introduction to mathematical techniques in pattern recognition. New York, NY, John Wiley, 1972.
- BERNSTEIN, E., ed. Digital image processing for remote sensing. New York, NY, IEEE Press, 1978.
- COMPUTER VISION GRAPHICS AND IMAGE PROCESSING. New York, Academic, 1972-bim.
- DEVIJVER, P.A.; KITTLER, J. Pattern Recognition: A Statistical Approach, Prentice Hall, 1982.

- DUDA, R.O.; HART, P.E. Pattern classification and scene analysis. New York, NY, John Wiley, 1973.
- FU, K.S.; ROSENFELD, A. Pattern recognition and image processing. IEEE Transactions on Computers. C-25(12):1336-1346, Dec. 1976.
- FUKUNAGA, K. Introduction to statistical pattern recognition. New York, NY, Academic.
- IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE. New York, NY, IEEE Computer Society, 1979- , bim.
- PATTERN RECOGNITION. Oxford, Pergamon, 1968- , bim.
- TOU, J.R.; GONZALEZ, R.C. Pattern recognition principles. Reading, MA., Addison Wesley, 1974.
- YOUNG, T.Y.; CALVERT, T.W. Classification, estimation and pattern recognition. New York, NY, American Elsevier, 1974.