

*Empresa Brasileira de Pesquisa Agropecuária
Embrapa Informática Agropecuária
Ministério da Agricultura, Pecuária e Abastecimento*

PROCESSAMENTO DE imagens digitais

MÉTODOS MULTIVARIADOS EM JAVA

José Iguelmar Miranda

*Embrapa Informática Agropecuária
Campinas, SP
2011*

Exemplares desta publicação podem ser adquiridos na:

Embrapa Informática Agropecuária

Av. André Tosello, 209 - Barão Geraldo

Caixa Postal 6041 - 13083-886 - Campinas, SP

Fone: (19) 3211-5700 - Fax: (19) 3211-5754

www.cnptia.embrapa.br

sac@cnptia.embrapa.br

Comitê de Publicações

Presidente: *Silvia Maria Fonseca Silveira Massruhá*

Membros: *Poliana Fernanda Giachetto, Roberto Hiroshi Higa, Stanley Robson de Medeiros Oliveira, Maria Goretti Gurgel Praxedes Adriana Farah Gonzalez, Neide Makiko Furukawa, Carla Cristiane Osawa*

Membros suplentes: *Alexandre de Castro, Fernando Attique Máximo, Paula Regina Kuser Falcão*

Secretária: *Carla Cristiane Osawa*

Supervisão editorial: *Neide Makiko Furukawa*

Revisão de texto: *Adriana Farah Gonzalez*

Normalização bibliográfica: *Marcia Izabel Fugisawa Souza*

Projeto gráfico/Capa/Editoração eletrônica: *Neide Makiko Furukawa*

Imagem da capa: *Embrapa Informática Agropecuária*

1ª edição 2011

1ª impressão (2011): tiragem 1000 exemplares

Todos os direitos reservados.

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei no 9.610).

**Dados Internacionais de Catalogação na Publicação (CIP)
Embrapa Informática Agropecuária**

Miranda, José Iguelmar.

Processamento de imagens digitais : métodos multivariados em Java / José Iguelmar Miranda. - Campinas, SP : Embrapa Informática Agropecuária, 2011.

400 p. : il.

ISBN 978-85-86168-02-4

1. Imagem digital. 2. Processamento de imagem. 3. Linguagem Java. 4. Estatística multivariada. 5. Redes de neurônios artificiais. 6. Lógica difusa. I. Título.

CDD - 21st ed

006.42

621.367

© Embrapa 2011

Agradecimento

O autor agradece à
Embrapa Informática Agropecuária
pelo apoio na publicação deste livro e à colega
Neide Makiko Furukawa pela dedicação
na elaboração e arte final do texto da obra.

Apresentação

O processamento de imagens digitais de satélite tornou-se uma ferramenta essencial de carácter estratégico. Graças a este processamento, pode-se hoje em dia estimar safras, contar pés de fruta em parcelas, avaliar o estado de estresse de uma determinada cultura agrícola, determinar faixas de bloqueio para evitar a dispersão de pragas e doenças, e monitorar deslocamentos de tropas em campos de batalha, dentre tantas outras aplicações.

A quantidade de satélites que nos disponibilizam imagens, incluindo os sino-brasileiros CBERS, também tem aumentado bastante. Via de regra, as imagens de satélite são disponibilizadas nas várias bandas espectrais em que seus sensores operam. Se, por um lado, estas várias informações sobrepostas nos permitem um leque maior de aplicações, por outro lado, requerem um ferramental matemático que considere as interdependências entre as imagens obtidas pelos diferentes sensores.

A explicação de como funciona a teoria matemática associada às análises multiespectrais pelos métodos multivariados é de grande interesse para todos os que se iniciam no assunto. O autor teve ainda o cuidado de rerepresentar vários conceitos da matemática do segundo grau para facilitar o aprendizado, como a teoria de matrizes, autovalores e autovetores, e permitindo uma leitura linear bastante fluida. Tem-se como resultado uma transição bem consolidada da base matemática com os novos conceitos do processamento digital de imagens.

O autor fornece ainda, gratuitamente, o código fonte, na linguagem de programação Java, implementando cada técnica, facilitando o uso em aplicações re-

ais que o leitor deseje fazer, seja simplesmente para entender o conteúdo, seja para as aplicações adaptadas à sua realidade prática. Os exemplos práticos e a possibilidade da sua execução computacional no ambiente do usuário facilitam em muito a aprendizagem, favorecendo o seu uso como livro didático.

Kleber Xavier Sampaio de Souza

Chefe-Geral

Embrapa Informática Agropecuária

Prefácio

A tendência atual, em termos de obtenção de dados de sensoriamento remoto, é a de que imagens com resoluções espaciais métricas e quilométricas sejam disponibilizadas gratuitamente na rede mundial de computadores. Apenas para citar um caso, cenas do território brasileiro do satélite norte-americano Landsat podem ser obtidas gratuitamente da página eletrônica do Instituto Nacional de Pesquisas Espaciais, mediante cadastro eletrônico. Obtenção de dados de sensoriamento remoto está deixando de ser um problema a ser resolvido.

As técnicas de processamento e análise de dados de sensoriamento remoto também têm evoluído significativamente. Na década de 1970, predominava a análise visual de imagens. O uso de lápis dermatográficos de diferentes cores para diferenciar redes de drenagens e sistemas viários em papéis transparentes sobrepostos às imagens analógicas era um procedimento bastante familiar na época. Somente na década de 1980, instituições governamentais e empresas privadas começaram a investir em pacotes computacionais para processamento de imagens. Nestes pacotes, predominavam algoritmos que não exigiam tempo computacional ou capacidade de armazenamento muito grande. A moda era o uso de técnicas de realce e de classificação relativamente simples como ampliação linear de contraste e classificação supervisionada pelo método do paralelepípedo.

A tendência atual é o uso de processamentos cada vez mais complexos como redes neurais e análises orientadas a objetos. Dados espectrais são combinados com dados auxiliares provenientes de mapas temáticos e cadastros para a tomada de decisões. Neste sentido, métodos multivariados são cada

vez mais utilizados, exigindo a incorporação de novas funções nos pacotes computacionais existentes. Quem domina uma linguagem de programação, tem grande chance de não se tornar escravo de softwares de processamento de imagens. O livro intitulado *Processamento de Imagens Digitais: Métodos Multivariados em Java*, escrito pelo pesquisador Dr. José Iguelmar Miranda, traz bons ventos àqueles que atuam na área de sensoriamento remoto. Isto porque, além de apresentar conceitos teóricos que são fundamentais no entendimento da análise multivariada, o autor disponibiliza ainda códigos-fonte em Java, uma linguagem que se destaca pela portabilidade, robustez e por ser orientada a objetos.

Edson Eyji Sano

Pesquisador da Embrapa Cerrados

Sumário

Parte 1	Motivação	19
Capítulo 1	Métodos multivariados.....	23
1.1	Introdução	23
1.2	Técnicas multivariadas.....	27
1.3	Organizando os dados	30
1.4	Onde encontrar o código fonte.....	32
Capítulo 2	Classes Java usadas no livro	33
2.1	Programação orientada a objetos e a linguagem Java.....	33
2.2	Classe e objetos	35
2.3	Subclasses e herança	42
2.4	Encapsulamento e atributos escondidos.....	45
2.5	Classes abstratas e interfaces	46
2.5.1	Classes abstratas.....	46
2.5.2	Interfaces	48
2.6	Classe e objeto na prática - criando a classe Imagem.....	52
2.7	APIs Java usadas	58
2.7.1	API JfreeChart.....	58
2.7.2	API GeoTools.....	59
2.7.3	API de álgebra linear - JAMA.....	59
2.7.4	API ANNeF	60

Parte 2 Fundamentos 61

Capítulo 3 - Álgebra de matrizes 65

3.1	Vetores.....	65
3.2	Matrizes.....	70
3.3	Tipos de matrizes	72
3.3.1	Matriz quadrada	72
3.3.2	Matriz nula	72
3.3.3	Matriz coluna	73
3.3.4	Matriz linha	73
3.3.5	Matriz diagonal	73
3.3.6	Matriz identidade	74
3.3.7	Matriz triangular superior	74
3.3.8	Matriz triangular inferior.....	74
3.3.9	Matriz simétrica	75
3.3.10	Traço de uma matriz.....	75
3.4	Operações com matrizes	75
3.4.1	Operações aritméticas de soma e subtração com matrizes..	75
3.4.2	Multiplicação de matriz por escalar	76
3.4.3	Transposta de uma matriz.....	77
3.4.4	Multiplicação de matrizes	78
3.4.5	Inversão de matrizes.....	79
3.5	Autovalores e autovetores de uma matriz quadrada	80

Capítulo 4 - Estatística univariada, bivariada e multivariada para imagens digitais 93

4.1	Introdução	93
4.2	Momentos de uma distribuição	93
4.3	Estatística de ordem	100
4.3.1	Mediana.....	101
4.3.2	Valores máximos e mínimos	102

4.3.3	Intervalo	102
4.3.4	Ponto médio	103
4.3.5	Desvio extremo	103
4.4	Estatística bivariada	104
4.4.1	Correlação linear de Pearson.....	104
4.4.2	Modelo de regressão linear	108
4.5	Regressão múltipla.....	116
4.6	Estatística descritiva de dados multivariados.....	123
4.7	Técnicas gráficas (Análise exploratória de dados).....	129

Capítulo 5 - Métricas de similaridade 133

5.1	Introdução	133
5.2	Medidas de similaridade	134
5.2.1	Medidas de correlação.....	136
5.2.2	Medidas de distância	136
5.2.3	Medidas de associação	142
5.3	Distância estatística.....	142
5.4	Outras distâncias	146
5.4.1	Distância de Mahalanobis	146
5.4.2	Distância de Minkowski.....	149
5.4.3	Distância de Bhattacharyya e seu uso na extração de atributos.....	149
5.4.4	Distância de Jeffreys-Matusita (JM)	151
5.5	Padronização dos dados	151
5.6	Resumo das métricas.....	152

Parte 3 Métodos Multivariados 155

Capítulo 6 - Análise de componentes principais 161

6.1	Introdução	161
-----	------------------	-----

6.2	Conceitos básicos de ACP.....	163
6.3	Componentes principais obtidas de variáveis uniformizadas	167
6.4	Aplicação de ACP no domínio espectral das imagens de satélite..	169
6.5	Aplicação de ACP no domínio temporal das imagens de satélite..	182

Capítulo 7 - Análise de agrupamentos 183

7.1	Introdução	183
7.2	O procedimento de agrupamento	186
7.2.1	Definir os objetivos da pesquisa.....	187
7.2.2	Fazer o projeto de pesquisa	187
7.2.3	Suposições na análise de agrupamento	188
7.2.4	Selecionando o algoritmo de agrupamento	189
7.2.5	Interpretação dos grupos	192
7.2.6	Validação dos grupos.....	193
7.3	Definições e notações.....	193
7.4	Técnicas de agrupamento.....	197
7.5	Algoritmos de particionamento de grupos - o método heurístico..	200

Capítulo 8 - Método k -médias 205

8.1	Introdução	205
8.2	Método de agrupamento do erro quadrático	208
8.3	O algoritmo k -médias.....	209
8.4	Exemplos e aplicações	215
8.5	Avaliando a precisão do mapa de classificação.....	219

Capítulo 9 - Método *fuzzy* c -médias (FCM) 225

9.1	Introdução	225
9.2	Teoria dos conjuntos difusos.....	228
9.3	Princípios para a aplicação da regra difusa.....	234
9.3.1	Difusividade	235

9.3.2	Inferência.....	241
9.3.3	De-difusividade (“Defuzzification”).....	244
9.4	Classificando com a regra difusa	247
9.5	O método <i>fuzzy c</i> -médias (FCM)	253
9.6	Classificação difusa: resumo	262
Capítulo 10 - Métodos da distância mínima da média e do paralelepípedo		269
10.1	Introdução aos métodos de classificação supervisionada	269
10.2	Esquema de classificação de cobertura/uso do solo	271
10.3	Seleção de áreas de treinamento	273
10.4	Observações práticas para a classificação supervisionada.....	276
10.4.1	Delimitação da área de treinamento	276
10.4.2	Extração das assinaturas espectrais	277
10.4.3	Classificação da imagem	281
10.5	O algoritmo de classificação da distância mínima da média	282
10.6	O método do paralelepípedo	294
Capítulo 11 - Método da máxima verossimilhança		301
11.1	Introdução	301
11.2	Axiomas da probabilidade	303
11.3	Teoria da decisão Bayesiana	307
11.4	O algoritmo maxver	309
11.5	Exemplo	315
11.6	A implementação do maxver.....	319
Capítulo 12 - Métodos usando redes de neurônios artificiais.....		323
12.1	Introdução	323
12.2	Perceptron - modelando um neurônio	325
12.3	Treinamento e aprendizado de uma rede de neurônios artificiais..	331

12.4	Perceptron multicamadas	348
12.5	O algoritmo da retropropagação	359
12.6	Aplicação	360
12.7	Outros modelos de redes de neurônios artificiais.....	380
12.7.1	Mapas auto-organizáveis de Kohonen - “ <i>self organizing maps</i> ” (SOM)	381
12.7.2	Rede de Hopfield.....	385
12.7.3	Teoria da ressonância adaptativa - “ <i>adaptive resonance theory</i> ” (ART)	387
	Referências	389
	Índice remissivo	395

Motivação