Reconhecimento de Padrões

Gonzalez - Capítulo 12

Hitoshi

2o Semestre 2004

Há muitas coisas que vemos todos os dias mas nunca nos damos conta que elas estão lá

Motivação

- ver x entender x perceber
- para que serve visão?
  - identificar regiões que tenham relevância para o agente visual, de forma remota
  - regiões = objetos ou padrões
- Padrão: arranjo de características (atributos) que podem ser extraídas da imagem

---

Como reconhecer padrões?

- Métodos estruturais
  - baseado em atributos qualitativos como ordem, posição relativa, etc.
- Teoria da decisão
  - baseado em atributos quantitativos como área, comprimento, etc

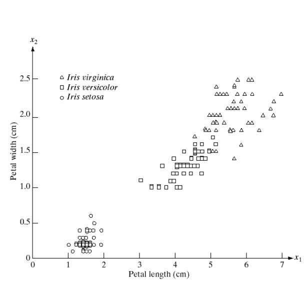
---

Definições

- Classes: ou classes de padrões são famílias de padrões definidas por w_1, w_2, ..., w_W, onde W é o número de classes
- Problema: atribuir um padrão (arranjo de característica extraídas da imagem) a uma classe, automaticamente
  - arranjo pode ser representado por um vetor, ou strings e árvores

---

FIGURE 12.1 Three types of iris flowers described by two measurements.



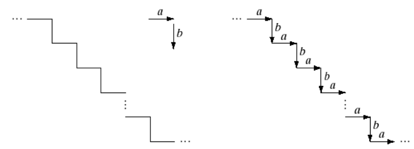
- Exemplo: classificação de flores
  - Há 3 flores (classes)
  - Cada flor é descrita por 2 atributos: comprimento e largura da pétala

A escolha dos atributos tem grande impacto no resultado da classificação

Atributos qualitativos

- Em alguns casos, relações estruturais são mais apropriados para descrever um padrão
  - impressões digitais
  - estrutura escada

Estrutura de escada



a b

FIGURE 12.3 (a) Staircase structure. (b) Structure coded in terms of the primitives ***a*** and ***b*** to yield the string description …***ababab***… .

seqüência de atributos, a e b, o string ababab torna o reconhecimento simples

Outro exemplo

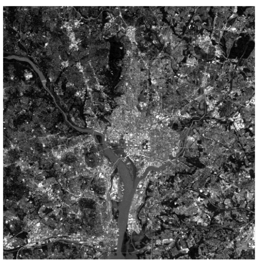


FIGURE 12.4 Satellite image of a heavily built downtown area (Washington, D.C.) and surrounding residential areas. (Courtesy of NASA.)

estrutura hierárquica

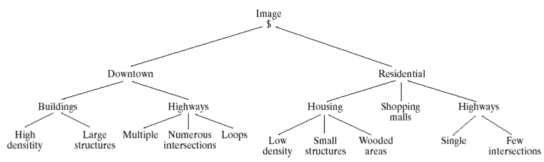


FIGURE 12.5 A tree description of the image in Fig. 12.4.

Teoria da decisão

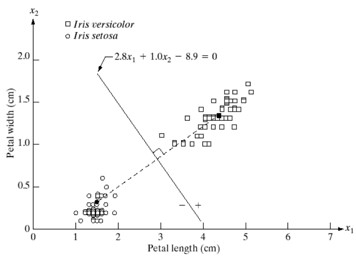


FIGURE 12.6 Decision boundary of minimum distance classifier for the classes of *Iris versicolor* and *Iris setosa*. The dark dot and square are the means.

Rec. Padrões - Teoria da Decisão

- As técnicas baseadas em teoria da decisão se utilizam de funções de decisão (funções discriminantes)
- Definições:
  - Vetor de padrões com n dimensões
    $\underline{\mathbf{x}} = (x_1, x_2, \ldots, x_n)^T$
  - Número de classes W
  - funções de decisão $d_1(\underline{\mathbf{x}})$, $d_2(\underline{\mathbf{x}})$, …, $d_W(\underline{\mathbf{x}})$

Definição do problema

- O problema de classificação tem solução se para um padrão $\underline{\mathbf{x}}$ pertencente à classe k

  $d_k(\underline{\mathbf{x}}) > d_l(\underline{\mathbf{x}})$, l=1,2, ...,W, k diferente de l
- Fronteira de decisão
  - separa duas classes wk e wl
  - é definida pelos valores de $\underline{\mathbf{x}}$ tal que

    $d_k(\underline{\mathbf{x}}) = d_l(\underline{\mathbf{x}})$, ou $d_{kl}(\underline{\mathbf{x}}) = d_k(\underline{\mathbf{x}}) - d_l(\underline{\mathbf{x}}) = 0$

Função de decisão

- Seja a função $d_{kl}(\underline{\mathbf{x}}) = d_k(\underline{\mathbf{x}}) - d_l(\underline{\mathbf{x}}) = 0$
  - que corresponde à fronteira de decisão entre as classes k e l
  - quando $d_{kl}(\underline{\mathbf{x}}) > 0$ temos que o padrão $\underline{\mathbf{x}}$ pertence à classe K
  - quando $d_{kl}(\underline{\mathbf{x}}) < 0$ temos que o padrão $\underline{\mathbf{x}}$ pertence à classe l

Correspondência (matching)

- cada classe é representada através de um vetor típico (característico)
- um vetor desconhecido é associado à classe mais próxima através de uma métrica pré-definida

Classificador de mínima distância

- o CMD é o classificador mais simples, baseado na distância euclidiana entre os vetores de padrão
- considere que cada classe já tenha sido definida (vetor típico tenha sido calculado através da média dos vetores usados para "aprende-la" (treiná-la)

$$\overline{m}_j = \frac{1}{N_j} \sum_{\overline{x} \in w_j} \overline{x}$$

p/ j=1,2,...,W
onde W = # de classes
N_j = # padrões de w_j

Treinamento x classificação

- Treinamento:
  - amostras de cada classe devem ser adquiridas para o cálculo dos $\underline{\mathbf{m}}_j$
- Classificação
  - um padrão desconhecido $\underline{\mathbf{x}}$ é associado à classe mais próxima, usando como critério de proximidade a distância euclidiana

    $D_j(\underline{\mathbf{x}}) = \| \underline{\mathbf{x}} - \underline{\mathbf{m}}_j \|$ j = 1,2,...,W

    onde $\| \underline{\mathbf{a}} \| = (\underline{\mathbf{a}}^T \underline{\mathbf{a}})^{1/2}$ => norma euclidiana

Função de decisão do CMD

Seja $d_k(\underline{\mathbf{x}}) = \underline{\mathbf{x}}^T \underline{\mathbf{m}}_k - 0.5 \underline{\mathbf{m}}_k^T \underline{\mathbf{m}}_k$ k=1,2,...,W
e a função de decisão será $d_{kl}(\underline{\mathbf{x}}) = d_k(\underline{\mathbf{x}}) - d_l(\underline{\mathbf{x}}) =$
$= \underline{\mathbf{x}}^T(\underline{\mathbf{m}}_k - \underline{\mathbf{m}}_l) - 0.5(\underline{\mathbf{m}}_k - \underline{\mathbf{m}}_l)^T(\underline{\mathbf{m}}_k + \underline{\mathbf{m}}_l) = 0$

- a função de decisão, definida por essa equação, é perpendicular ao segmento de linha entre $\underline{\mathbf{m}}_k$ e $\underline{\mathbf{m}}_l$.
  - quando n=2 a superfície é uma reta
  - quando n=3 um plano
  - quando n>3 um hiperplano

Exemplo: íris

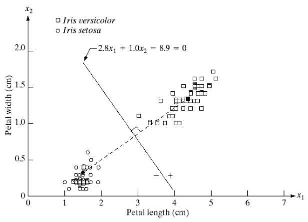
Sejam:
$m_1 = (4.3, 1.3)^T$ e
$m_2 = (1.5, 0.3)^T$

Decisão:

$d_1(\underline{\mathbf{x}}) = \underline{\mathbf{x}}^T\underline{\mathbf{m}}_1 - 0.5\underline{\mathbf{m}}_1^T\underline{\mathbf{m}}_1 =$
$4.3x_1 + 1.3x_2 - 10.1$

$d_2(\underline{\mathbf{x}}) = \underline{\mathbf{x}}^T\underline{\mathbf{m}}_2 - 0.5\underline{\mathbf{m}}_2^T\underline{\mathbf{m}}_2 =$
$1.5x_1 + 0.3x_2 - 1.17$

$d_{12}(\underline{\mathbf{x}}) = d_1(\underline{\mathbf{x}}) - d_2(\underline{\mathbf{x}}) =$
$2.8x_1 + 1.0x_2 - 8.9$



Na prática, o CMD funciona bem quando a distância entre as médias é grande em comparação com a variância das classes

Correspondência por correlação

- dada uma subimagem w(x,y) de tamanho JxK de uma imagem f(x,y) de tamanho MxN (J<=M, K<=N) a correlação entre f(x,y) e w(x,y) é definida por:

$$c(x,y) = \sum_s \sum_t f(s,t)\,w(x+s, y+t)$$

$$x = 0,1,2,\ldots, M-1$$

$$y = 0,1,2,\ldots, N-1$$

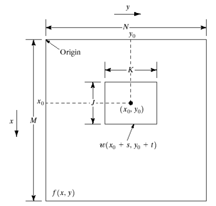


FIGURE 12.8 Arrangement for obtaining the correlation of f and w at point (x_0, y_0).

Observações:
- funções reais
- constante eliminada por normalização

Cuidados

- o valor máximo de c(x,y) indica os pontos de melhor correspondência
- bordas da imagem continuam um problema
- sensível a alterações de intensidade:
  - dobrando f(x,y) -> dobra c(x,y)

Coeficiente de correlação

$$\gamma(x,y) = \frac{\sum_s \sum_t [f(s,t) - \bar{f}(s,t)][w(x+s,y+t) - \bar{w}]}{\{\sum_s \sum_t [f(s,t) - \bar{f}(s,t)]^2 \sum_s \sum_t [w(x+s,y+t) - \bar{w}]^2\}^{1/2}}$$

$\bar{w}$ = média dos pixels de w

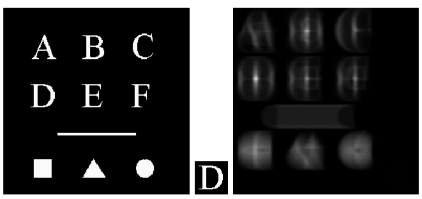
$\bar{f}$ = média dos pixels de f na região w

s, t = região comum a w e t

$0 \le x \le M-1$ e $0 \le y \le N-1$

O coeficiente $\gamma(x,y)$ permanece no intervalo [-1,1] independentemente das mudanças de amplitudes de w e f

Exemplo



a b c
FIGURE 12.9 (a) Image. (b) Subimage. (c) Correlation coefficient of (a) and (b). Note that the highest (brighter) point in (c) occurs when subimage (b) is coincident with the letter "D" in (a).

Correlação

- tolera mudanças de intensidade usando
- é mais difícil tratar mudanças de
  - escala
  - rotações
- computacionalmente caro
  - inviável para escala/rotação (atualmente)
  - FFT melhor quando tamanho de W > 13x13

Classificadores estatísticos ótimos

- Na interpretação de muitos eventos naturais é importante considerar incertezas inerentes às classes através do uso de probabilidades.
- Dessa forma é possível desenvolver um método de classificação ótimo
  - ótimo = na média, seus resultados correspondem à menor probabilidade de se cometer erros de classificação

Fundamentos

- probabilidade condicional de uma padrão x pertencer à classe wi é dado por: p(wi | x)
- se o classificador decide que o padrão x veio na verdade de uma classe wj quando na verdade é da classe wi, ele incorre uma perda dada por Lij
- x pode pertencer, a princípio, a qualquer classe wj, ou seja, a perda média pode ser calculada por:

$$r_j(\mathbf{x}) = \sum_{k=1}^{W} L_{kj}\ p(w_k \mid \mathbf{x})$$

conhecido como perda, risco, ou erro condicional médio

Fundamentos

- sabemos que p(A|B) = p(B|A)p(A)/p(B)
- e assim:

$$r_j(\mathbf{x}) = \frac{1}{p(\mathbf{x})} \sum_{k=1}^{W} L_{kj}\ p(\mathbf{x} \mid w_k) P(w_k)$$

$P(w_k) =$ prob. de ocorrência de w_k

$p(\mathbf{x} \mid w_k) =$ fdp dos padrões de w_k

$p(\mathbf{x}) =$ positivo e comum a todos r_j e portanto pode ser ignorado

$$r_j(\mathbf{x}) = \sum_{k=1}^{W} L_{kj}\ p(\mathbf{x} \mid w_k) P(w_k)$$

Classificadores bayesianos

- Dado um padrão x, o classificador tem W classes possíveis para escolher
- Para se escolher a classe, calcula-se as perdas r1(x), r2(x), .., rw(x)
- O padrão x é associado à classe de menor risco.
- Tais classificadores são conhecidos como bayesianos

$$r_i(\mathbf{x}) < r_j(\mathbf{x}) \text{ para todo j, j} \neq \text{i}$$

$$\sum_{k=1}^{W} L_{ki}\ p(\mathbf{x} \mid w_k) P(w_k) < \sum_{q=1}^{W} L_{qi}\ p(\mathbf{x} \mid w_q) P(w_q)$$

Perda (Custo)

- Seja
  - 0 = custo (perda) de uma decisão correta
  - 1 = custo de uma decisão incorreta
- assim Lij = 1 - δij, onde
  - δij =0 se i<>j
  - δij =1 se i=j

$$r_j(\mathbf{x}) = \sum_{k=1}^{W} (1-\partial_{kj})\, p(\mathbf{x} \mid w_k) P(w_k) = p(\mathbf{x}) - p(\mathbf{x} \mid w_j) P(w_j)$$

Classificador bayesiano

- associa a classe wi ao padrao x se, para todo j diferente de i, se

$$p(\mathbf{x}) - p(\mathbf{x} \mid w_i) P(w_i) < p(\mathbf{x}) - p(\mathbf{x} \mid w_j) P(w_j)$$

$$\text{ou} \quad p(\mathbf{x} \mid w_i) P(w_i) > p(\mathbf{x} \mid w_j) P(w_j), \ \forall \text{i} \neq \text{j}$$

- Observe que, de nossa discussão inicial, p(x|wj)P(wj) pode ser considerado uma função de decisão dj(x), onde a classe é associada ao maior valor dessa função

Correlação

- a função de decisão

$$c(x, y) = \sum_s \sum_t f(s,t)w(x+s, y+t)$$

também é ótima no sentido de minimizar o risco de erros na classificação. Porém, é necessário conhecer:
  - a) a pdf dos padrões de cada classe p(x|wi)
  - b) a prob. de ocorrência de cada classe P(wi)

Modelos probabilísticos

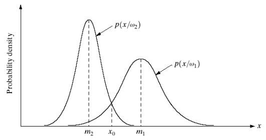
- Se todas as classes tiverem a mesma probabilidade, P(wi) pode ser assumida como 1/W, onde W é o número total de classes
  - mesmo que isso não seja verdade, em geral é possível obter P(wi)
- estimar p(x|wi) é mais difícil. Por isso é comum assumir que essas pdf's seguem alguma expressão analítica
  - amostras dos padrões são utilizadas para estimar os parâmetros dessas expressões
  - quanto mais próximas da realidade esses modelos probabilísticos, mais preciso será o classificador

Caso gaussiano 1D

- considere o caso 1D (n=1) em que dispomos de 2 classes (W=2), com pdf's gaussianas, com médias m1 e m2 e desvios padrão s1 e s2, respectivamente.
- A função de decisão tem a forma

$$d_j(x) = p(x \mid w_j)P(w_j) = \frac{1}{\sqrt{2\pi}s_j} e^{\frac{-(x-m_j)^2}{2s_j^2}} P(w_j)$$

FIGURE 12.10 Probability density functions for two 1-D pattern classes. The point x_0 shown is the decision boundary if the two classes are equally likely to occur.



Caso gaussiano n-dimensional

- A pdf dos vetores da classe j tem a forma:

$$p(\mathbf{x} \mid w_j) = \frac{1}{(2\pi)^{n/2} |\mathbf{C}j|^{1/2}} e^{-(\mathbf{x}-\mathbf{m}_j)^T \mathbf{C}j^{-1}(\mathbf{x}-\mathbf{m}_j)/2}$$

$$\mathbf{C}_j = \text{matriz de covariância} = \mathrm{E}_j\{(\mathbf{x}-\mathbf{m}_j)(\mathbf{x}-\mathbf{m}_j)^T\}$$

$$|\mathbf{C}_j| = \text{determinante de } \mathbf{C}_j$$

$$\mathbf{m}_j = E_j\{\mathrm{x}\}$$

- onde n é o número de atributos em um padrão
- Ej{.} = valor esperado do argumento sobre os padrões da classe j

Valor médio

- se o valor esperado for aproximado pela média dos padrões então:

$$\mathbf{m}_j = \frac{1}{Nj} \sum_{\mathbf{x} \in w_j} \mathbf{x} \quad \text{e} \quad \mathbf{C}\mathrm{j} = \frac{1}{Nj} \sum_{\mathbf{x} \in w_j} \mathbf{x}\mathbf{x}^T - \mathbf{m}_j\mathbf{m}_j^T$$

onde Nj é o número de padrões de w_j

Propriedades da matriz de covariância

- é simétrica e semi definida positiva.
- um elemento da diagonal principal (Ckk) corresponde a variância do k-ésimo elemento do padrão
- um elemento fora da diagonal principal (Cjk) corresponde a covariância entre xj e xk.
- quando os elementos não são correlacionados
  - Cjk = 0 para j<>k, e o problema multivariado é reduzido a um problema univariado

Função de decisão

- devido à exponencial da pdf gaussiana, é mais natural considerar o log da função de decisão bayesiana

  $d_j(x)=p(x|w_j)P(w_j)=\ln[p(x|w_j)]+\ln[P(w_j)]$

- isso não afeta o resultado da decisão pois o log é uma função de crescimento monotonico, e portanto a ordem dos valores usados para tomada de decisão é mantida

$d_j(x) = \ln[P(w_j)] - (n/2)\ln(2\pi) - (1/2)\ln(|C_j|)$
$\quad - (1/2)[(x-m_j)^T C_j^{-1}(x-m_j)]$

função de decisão

- eliminando o termo $(n/2)\ln(2\pi)$ por ser cte:

$d_j(x)= \ln[P(w_j)] - (1/2)\ln(|C_j|) - (1/2)[(x-m_j)^T C_j^{-1}(x-m_j)]$

- essa equação corresponde a hiperquádricas (funções quadráticas no espaço n-dimensional)
  - ou seja, as superfícies de decisão entre pares de classes são funções de 2o grau
  - qdo as matrizs de covariância são todas iguais, Cj=C para todo j

    $d_j(x)= \ln[P(w_j)] + x^T C^{-1} m_j] - (1/2) m_j^T C^{-1} m_j$

    que correspondem a hiperplanos

Função de decisão

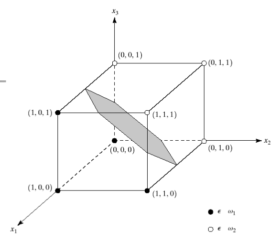
- Se C = I (matriz identidade), então P(wj) = 1/W

  $d_j(x)= x^T m_j - (1/2) m_j^T m_j$

- que equivale as funções de decisão de um classificador de mínima distância
- Portanto um CMD é ótimo no sentido bayesiano quando:
  - 1) a pdf dos padrões das classes são gaussianos
  - 2) matrizes de covariâncias = I
  - 3) classes são equiprováveis

Isso implica que as classes formam nuvens esféricas em n dimensões (hiper esferas). o CMD define hiperplanos perpendiculares à linha que liga os dois centros

Exemplo

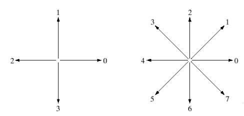


- Considere a figura acima, que mostra a distribuição de 2 classes em 3D. Calcule a função de decisão para um classificador Bayesiano.

Métodos estruturais

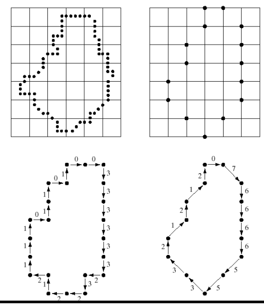
- além de utilizar apenas atributos quantitativos, podemos utilizar relações de estrutura para reconhecer padrões.

- Como representar e descrever estruturas?



a b
FIGURE 11.1 Direction numbers for (a) 4-directional chain code, and (b) 8-directional chain code.

- Cadeias
  - vizinhança-4
  - vizinhança-8
- podem ser muito longas
  - re-amostragem
- para não depender do ponto inicial, tratar a cadeia como circular
- rotações podem ser compensadas usando a 1a diferença da cadeia: número de mundanças de direção no sentido anti-horário



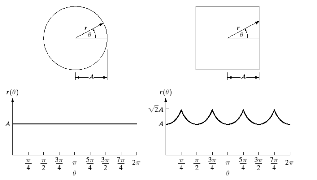
a b
c d
FIGURE 11.2 (a) Digital boundary with resampling grid superimposed. (b) Result of resampling. (c) 4-directional chain code. (d) 8-directional chain code.

Assinaturas

- Assinaturas são representações unidimensionais de um contorno, como a variação do raio com relação ao baricentro
  - são invariantes a translação, mas dependem da escala e rotação
  - a rotação pode ser compensada se existir um ponto característico (ex.: ponto mais distante)

a b

FIGURE 11.5 Distance-versus-angle signatures. In (a) $r(\theta)$ is constant. In (b), the signature consists of repetitions of the pattern $r(\theta) = A \sec\theta$ for $0 \le \theta \le \pi/4$ and $r(\theta) = A \csc\theta$ for $\pi/4 < \theta \le \pi/2$.



Descritores simples

- Comprimento: número de pixels pertencentes a borda (perímetro)
- Diâmetro: max[D(pi,pj)]
  - pi, pj, pontos sobre a borda
  - define o eixo principal
  - eixo secundário: perpendicular ao eixo principal, de comprimento tal que contenha a borda, definindo o retângulo básico
  - excentricidade: razão entre os comprimentos dos eixos principal e secundário
- curvatura: taxa de mudança da inclinação

Número de forma (shape number)

- vimos que a 1a diferença pode ser usada para representar um contorno, mas a cadeia correspondente depende do ponto inicial.
- Número de forma:
  - número de menor magnitude, possível de ser obtida a partir de rotações da cadeia formada pelas 1a diferenças
- a ordem n do número de forma é definido como o número de dígitos utilizados na representação

Exemplo número de forma

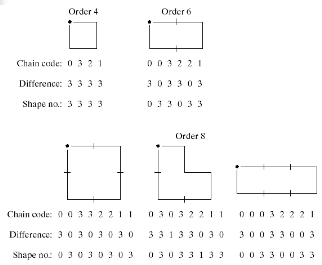


FIGURE 11.11 All shapes of order 4, 6, and 8. The directions are from Fig. 11.1(a), and the dot indicates the starting point.

note que a ordem é sempre par para contornos fechados

Reconhecimento de formas

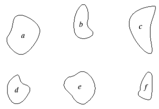
- O número de forma pode ser utilizado para reconhecer contornos.
  - o grau de similaridade (k) entre dois contornos pode ser definido como a maior ordem sob a qual o número de forma continua o mesmo, ou seja:
    - sj(a) = sj(b) para j=4,6,...,k
    - sj(a) <> sj(b) para j=k+2,k+6,...
    - onde s corresponde ao número de forma dos contornos a e b

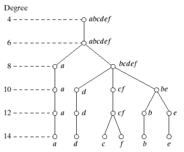
Correspondência usando número de forma

- a distância entre 2 formas a e b pode ser definida como o inverso da similaridade: D(a,b) = 1/k
  - satisfaz as seguintes propriedades:
    - D(a,b) >= 0
    - D(a,b) = 0 iff a = b
    - D(a,c) <= max [D(a,b), D(b,c)]

exemplo

- dado a forma f, qual a forma mais próxima?





	a	b	c	d	e	f
a	∞	6	6	6	6	6
b		∞	8	8	10	8
c			∞	8	8	12
d				∞	8	8
e					∞	8
f						∞

a
b c
FIGURE 12.24 (a) Shapes. (b) Hypothetical similarity tree. (c) Similarity matrix. (Bribiesca and Guzman.)

Correspondência com strings

- assim como o contorno pode ser transformado em cadeias de direções, ele pode também ser codificado na forma de strings, onde cada elemento do string (letra) pode corresponder a uma parte do contorno
  - exemplo: escada

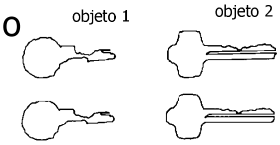
Similaridade entre strings

- considere 2 contornos *a* e *b*, codificados como os strings *a1a2...an* e *b1b2...bm*, respectivamente.
- Seja A o número de correspondências entre *a* e *b*
  - uma correspondência ocorre na k-ésima posição se ak = bk.
  - o número de símbolos que não tem correspondência é B = max(|a|,|b|) - A
  - |a| indica o comprimento do string a
  - pode-se mostrar que B = 0 iff a=b
- uma medida de similaridade possível
  - R = A/B = A / max(|a|,|b|) - A
  - R é infinito quando a correspondência é perfeita
  - R é zero quando não há nenhuma correspondência
  - ponto inicial é importante para reduzir custo computacional

exemplo

- os strings são formados a partir dos ângulos internos dos polígonos, visitados no sentido horário.
- Foram usados 8 símbolos, correpondentes a cada octante.

objeto 1 objeto 2



R	1.a	1.b	1.c	1.d	1.e	1.f
1.a	∞					
1.b	16.0	∞				
1.c	9.6	26.3	∞			
1.d	5.1	8.1	10.3	∞		
1.e	4.7	7.2	10.3	14.2	∞	
1.f	4.7	7.2	10.3	8.4	23.7	∞

R	2.a	2.b	2.c	2.d	2.e	2.f
2.a	∞					
2.b	33.5	∞				
2.c	4.8	5.8	∞			
2.d	3.6	4.2	19.3	∞		
2.e	2.8	3.3	9.2	18.3	∞	
2.f	2.6	3.0	7.7	13.5	27.0	∞

R	1.a	1.b	1.c	1.d	1.e	1.f
2.a	1.24	1.50	1.32	1.47	1.55	1.48
2.b	1.18	1.43	1.32	1.47	1.55	1.48
2.c	1.02	1.18	1.19	1.32	1.39	1.48
2.d	1.02	1.18	1.19	1.32	1.29	1.40
2.e	0.93	1.07	1.08	1.19	1.24	1.25
2.f	0.89	1.02	1.02	1.24	1.22	1.18

a b
c d
e f
g
FIGURE 12.25 (a) and (b) Sample boundaries of two different object classes; (c) and (d) their corresponding polygonal approximations; (e)–(g) tabulations of *R*. (Sze and Yang.)

Reconhecimento sintático

- fornece um formalismo para problemas de reconhecimento de estruturas
- Metodologia:
  - considere 2 classes w1 e w2
  - os padrões são strings formados por primitivas que podem ser obtidos da imagem. Essas primitivas podem ser considerados símbolos do alfabeto de uma gramática
  - uma gramática (G) é um conjunto de regras de sintaxe, utilizadas para a construção de sentenças
  - o conjunto de sentenças geradas por uma gramática G é denominada linguagem L(G)
  - portanto, sentenças são strings de símbolos (padrões), e as linguagens correspondem às classes

Gramáticas para reconhecer padrões

- dadas 2 gramáticas G1 e G2
  - G1 gera sentenças que correspondem a padrões da classe w1
  - G2 gera sentenças que correspondem a padrões da classe w2
- uma vez que essas gramáticas estejam defidas, o processo de classificação é simples:
  - se o string pertence a L(G1) então ele pertence a w1
  - de forma semelhante, se o string pertence a L(G2), então ele pertence a w2
  - a decisão não é possível se o string pertencer às duas linguagens
  - um string que não pertença a nenhuma linguagem é rejeitado

Gramáticas

- uma gramática é definida pela 4-tupla: G=(N, T, P, S)
  - N = conjunto de variáveis ou não terminais
  - T = conjunto finito de constantes ou terminais
  - P = conjunto de regras de produção
  - S = elemento de N, símbolo inicial, de onde são derivadas todas as sentenças de G

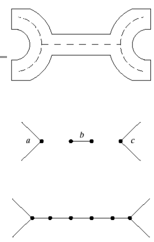
Notação

- elementos de N: descritos por letras maiúsculas
- elementos de T: letras minúsculas no começo do alfabeto
- letras minúsculas no final do alfabeto = strings de terminais
- letras gregas = strings com terminais e não terminais.
- λ = Sentença vazia
- V* = todas as sentenças compostas por elementos de V

Gramáticas regulares

- essas gramáticas tem regras de produção da forma:
  - A -> aB
  - A -> a
  - ou seja, regras da forma A->α, com A em N e α em (N $\cup$ T)*

Exemplo:



a
b
c

FIGURE 12.26 (a) Object represented by its (pruned) skeleton. (b) Primitives. (c) Structure generated by using a regular string grammar.

- G = {N, T, P, S}
  - N = {A, B, S}
  - T = {a, b, c}
  - P = {S->aA,
    A->bA,
    A->bB,
    B->c}

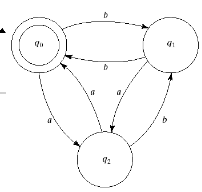
- Exemplifique as regras que geram abbbbbc
- Qual a linguagem gerada por essa gramática?

Reconhecedores de strings

- gramáticas servem para *gerar* padrões
- como resolver o problema de verificar se uma sentença pertence ou não à linguagem L(G)?
  - solução: autômatos finitos
  - autômatos finitos são definidos pela quintupla:
    A_{finito} = (Q, T, d, q_0, F)
    - Q = conjunto não vazio de estados
    - T = alfabeto
    - d é um mapeamento QxT sobre subconjuntos de Q
    - q_0 é o estado inicial
    - F são estados finais (ou de aceitação)

Exemplo

estado final



- Considere o autômato:
- onde:
  - Q = {q_0,q_1,q_2},
  - T={a,b}
  - F={q_0}
  - d(q0,a)={q_2}, d(q0,b)={q_1}, d(q1,a)={q_2}, d(q1,b)={q_0}, d(q2,a)={q_0}, d(q2,b)={q_1}

- Teste os strings: abbabb e aabab

gramáticas e autômatos

- Há uma correspondência única entre gramáticas regulares e autômatos finitos, ou seja, uma linguagem é reconhecida por um autômato finito iff ela é gerada por uma gramática regular
- é possível definir um autômato a partir da gramática
  - seja G = {N, T, P, X0}
  - onde N = {X0, X1, ... , Xn}
  - o conjunto Q do {q0, q1, ..., qn, qn+1}, tal que qi corresponde a Xi, e Xn+1 é o estado final.
  - o conjunto de terminais é o mesmo conjunto T de G

Gramática x autômato

- O mapeamento d(.,.) é obtido por:
  - if Xi ->aXj está em P, então d(qi,a)={qj}
  - if Xi-> a está em P, então d(qi,a)={qn+1}
- De forma semelhante, a partir de uma autômato finito é possível obter a gramática geradora, sendo N composto pelos elementos de Q, o símbolo X0 corresponde a q0, e as regras de produção:
  - if d(qi,a)={qj}, há uma regra Xi -> aXj em P
  - if d(qi,a)=um estado em F, existe Xi -> a em P

Exemplo

- G = {N, T, P, S}
  - N = {A, B, S}, T = {a, b, c}, P = {S->aA, A->bA, A->bB, B->c}
- encontre o autômato associado a G
- A partir do autômato, encontre a gramática G'.