

Modelos de Markov Não
Observáveis e Gramáticas
Estocásticas Regulares

Ana L. N. Fred

Instituto Superior Técnico



HMM

vs

SFSG:

(Hidden Markov Models)

(Stochastic Finite State Grammars)

Modelos de Markov Não Observáveis

Gramáticas Estocásticas de Estados Finitos

- São instâncias de uma classe mais geral de modelos: redes estocásticas de estados finitos (*Stochastic Finite-State Networks*)

- conj. finito de estados

- distribuições de probabilidade que definem transições entre estados e a produção de sequências finitas de observações

- baseados na teoria dos processos estocásticos, as suas origens são diferentes:

- teoria da informação - modelos de Markov

- extensões da teoria das linguagens formais

- gramáticas estocásticas regulares

- autômatos estocásticos de estados finitos

- Ambos geram uma sequência interna (não-observável) de símbolos (estados) e uma sequência externa (observável) de símbolos usando regras probabilísticas.

- Assumem formalismos diferentes e mecanismos distintos de inferência.

- A probabilidade de uma sequência é calculada de uma forma semelhante

Relações formais entre modelos no
contexto das linguagens geradas

Σ - alfabeto

$$\Sigma = \{a, b, c\}$$

aabbacbabbacc
aa
bba

Σ^* - concatenação de símbolos em Σ

$L \subseteq \Sigma^*$ - linguagem

$SL = (L, p)$ linguagem estocástica

$$L \subseteq \Sigma^*$$

$$p : \Sigma^* \rightarrow [0,1]$$

$$p(x) = 0 \quad \text{se } x \notin L$$

$$\sum_{x \in L} p(x) = 1$$

WL - *weighted language*

$SRL = (L, p)$ tal que L é regular

(= aceite por um autómato de estados finitos)

- linguagem estocástica regular

Modelos de Markov

HMM

$$H = (Q, \Sigma, A, B, \pi)$$

$Q = \{q_i\} \equiv$ conjunto finito de estados

$\Sigma \equiv$ conjunto de símbolos de observação

$A: Q \times Q \rightarrow [0,1] \equiv$ matriz de probabilidade de transição

$$\sum_{q' \in Q} A(q, q') = 1, \forall_{q \in Q}$$

$B: Q \times \Sigma \rightarrow [0,1] \equiv$ distribuição de probabilidade de observação de símbolo

$$\sum_{a \in \Sigma} B(q, a) = 1, \forall_{q \in Q}$$

$\pi: Q \rightarrow [0,1] \equiv$ distribuição de probabilidade do estado inicial

$$\sum_{q \in Q} \pi(q) = 1$$

Probabilidade de observação da sequência

$$x = x_1, x_2 \dots x_n, \quad x_i \in \Sigma$$

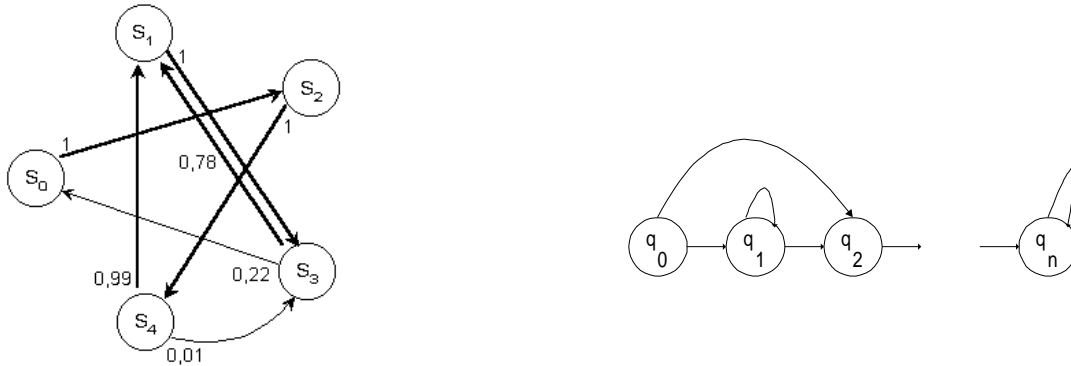
$$p(x | H) = p_H(x) = \sum_{q_1 \dots q_n} \pi(q_1) B(q_1, x_1) A(q_1, q_2) B(q_2, x_2) \dots A(q_{n-1}, q_n) B(q_n, x_n)$$

Proposição: Dado um HMM definido em Σ , p_H define uma distribuição de probabilidade em Σ^n para cada $n \in \mathbb{N}$.

$$\begin{aligned} & \sum_{x_1 \dots x_n \in \Sigma^n} p_H(x_1 \dots x_n) = \\ &= \sum_{x_1 \dots x_n} \sum_{q_1 \dots q_n} \pi(q_1) B(q_1, x_1) A(q_1, q_2) B(q_2, x_2) \dots A(q_{n-1}, q_n) B(q_n, x_n) \\ &= \sum_{q_1 \dots q_n} \pi(q_1) \left(\sum_{x_1} B(q_1, x_1) A(q_1, q_2) \left(\sum_{x_2} \dots \left(\sum_{x_n} B(q_n, x_n) \right) \dots \right) \right) \\ &= \sum_{q_1} \pi(q_1) \sum_{q_2} A(q_1, q_2) \left(\dots \left(\sum_{q_n} A(q_{n-1}, q_n) \right) \dots \right) \\ &= 1 \end{aligned}$$

HMM usando restrições temporais

LRHMM (*left-to-right HMM*)



LRHMM - modelo esquerda-direita

$$H_{LR} = (Q, \Sigma, A, B, q_i, q_f)$$

$Q = \{q_i\} \equiv$ conjunto finito de estados ordenados tal que
se $q < q'$ então $A(q', q) = 0$

$q_i \equiv$ estado inicial

$q_f \equiv$ estado final

H_{LR} define uma função $p_{HLR} : \Sigma^* \rightarrow [0,1]$

Probabilidade de observação da sequência

$$x = x_1, x_2 \dots x_n, \quad x_i \in \Sigma$$

$$p_{HLR}(x) = \sum_{q_2 \dots q_{n-1}} B(q_i, x_1) A(q_i, q_2) B(q_2, x_2) \dots A(q_{n-1}, q_f) B(q_f, x_n)$$

A introdução do conceito de estado final modifica as propriedades de geração estatística de strings

Proposição: $\Gamma_{HMM} \cap \Gamma_{HLR} \neq \emptyset$ $\Gamma_{HMM} \not\subset \Gamma_{HLR}$
 $\Gamma_{HLR} \not\subset \Gamma_{HMM}$

⋮
HMMT - *HMM with observation probability distribution in the transitions*
 ⋮

HMMT - definição alternativa (equivalente) de um HMM em que a distribuição de probabilidade de observação de símbolo é atribuída às transições em vez de aos estados.

$$\begin{aligned}
 HT &= (Q, \Sigma, A, B, \pi) \\
 Q, A, \pi &\text{ como definidos anteriormente} \\
 B: Q \times \Sigma \times Q &\rightarrow [0,1] \\
 \sum_{a \in \Sigma} B(q, a, q') &= 1 \quad \forall_{q, q' \in Q}, \text{ se } A(q, q') \neq 0
 \end{aligned}$$

Probabilidade de observação da sequência

$$x = x_1, x_2 \dots x_n, \quad x_i \in \Sigma$$

$$p_{HT}(x) = \sum_{q_0 \dots q_n} \pi(q_0) A(q_0, q_1) B(q_0, x_1, q_1) \dots A(q_{n-1}, q_n) B(q_{n-1}, x_n, q_n)$$

Proposição:

Para cada HMM $H = (Q, \Sigma, A, B, \pi)$ existe um HMMT $HT = (Q', \Sigma, A', B', \pi')$ tal que $\forall_{x \in \Sigma^*}, p_H(x) = p_{HT}(x)$

Proposição:

Para cada HMMT $HT = (Q, \Sigma, A, B, \pi)$ existe um HMM $H = (Q', \Sigma, A', B', \pi')$ tal que $\forall_{x \in \Sigma^*}, p_H(x) = p_{HT}(x)$

Proposição:

Para cada HMM $H = (Q, \Sigma, A, B, \pi)$ existe um HMMT $HT = (Q', \Sigma, A', B', \pi')$ tal que $\forall_{x \in \Sigma^*}, p_H(x) = p_{HT}(x)$

Demonstração:

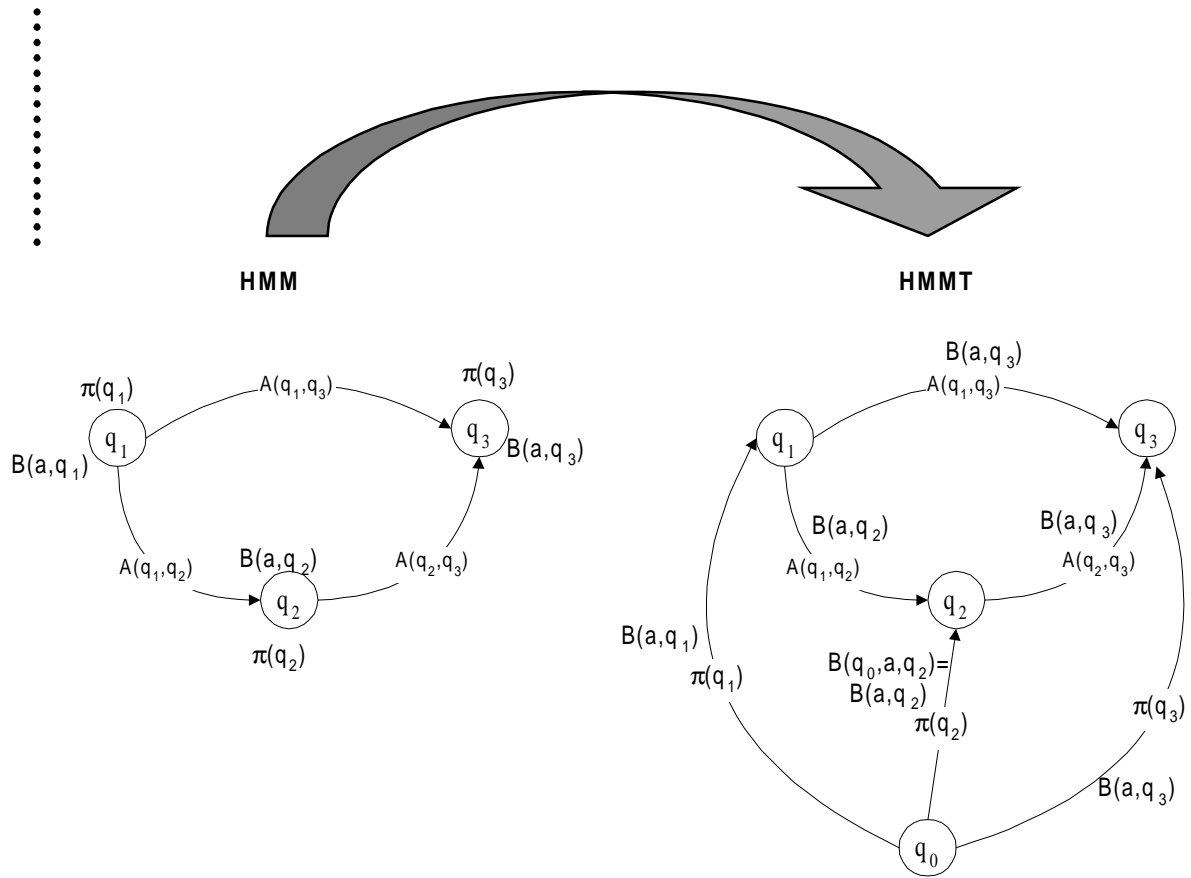
$$\begin{aligned} \text{Seja } Q' &= Q \cup \{q_0\} \text{ com } q_0 \notin Q \\ A'(q, q') &= A(q, q') \quad \forall q, q' \in Q \\ A'(q_0, q) &= \pi(q) \quad \forall q \in Q \\ B'(q, a, q') &= B(a, q') \quad \forall a \in \Sigma, \forall q, q' \in Q \\ B'(q_0, a, q) &= B(a, q) \quad \forall a \in \Sigma, \forall q \in Q \\ \pi'(q_0) &= 1 \quad \text{e } \pi'(q) = 0 \quad \forall q \in Q \end{aligned}$$

a) As novas distribuições verificam a definição de HMMT:

$$\begin{aligned} \sum_{q' \in Q} A'(q, q') &= \sum_{q' \in Q} A(q, q') = 1 \quad \forall q \in Q \\ \sum_{q \in Q} A'(q_0, q) &= \sum_{q \in Q} \pi(q) = 1 \\ \sum_{a \in \Sigma} B'(q, a, q') &= \sum_{a \in \Sigma} B(a, q') = 1 \quad \forall q' \in Q \\ \sum_{a \in \Sigma} B'(q_0, a, q) &= \sum_{a \in \Sigma} B(a, q) = 1 \quad \forall q \in Q \end{aligned}$$

b) A equivalência é mostrada por: para todo o x com $|x|=n$

$$\begin{aligned} p_{HT}(x) &= \sum_{q_0 \dots q_n} \pi'(q_0) A'(q_0, q_1) B'(q_0, x_1, q_1) \\ &\quad \dots A'(q_{n-1}, q_n) B'(q_{n-1}, x_n, q_n) \\ &= \sum_{q_1 \dots q_n} \pi(q_1) B(x_1, q_1) A(q_1, q_2) \\ &\quad \dots A(q_{n-1}, q_n) B(x_n, q_n) \\ &= p_H(x) \end{aligned}$$



Proposição:

Para cada HMMT $HT = (Q, \Sigma, A, B, \pi)$ existe um HMM $H = (Q', \Sigma, A', B', \pi')$ tal que $\forall_{x \in \Sigma^*}, p_H(x) = p_{HT}(x)$

Demonstração:

Seja

$$Q' = \{(q, q') \mid q, q' \in Q \text{ e } A(q, q') \neq 0\}$$

$$A'((q, q'), (q'', q''')) = A(q'', q''') \quad \text{se } q' = q''$$

$$= 0 \quad \text{caso contrário}$$

$$B'(a, (q, q')) = B(q, a, q')$$

$$\pi'((q, q')) = \pi(q)A(q, q')$$

a) As novas distribuições verificam a definição de HMM:

$$\sum_{(q'', q''')} A'((q, q'), (q'', q''')) = \sum_{q''} A(q', q'') = 1 \quad \forall q, q' \in Q$$

$$\sum_{a \in \Sigma} B'(a, (q, q')) = \sum_{a \in \Sigma} B(q, a, q') = 1 \quad \forall q, q' \in Q$$

$$\sum_{(q, q')} \pi'((q, q')) = \sum_{q, q'} \pi(q)A(q, q') = 1$$

b) A equivalência é mostrada por: para todo o x com $|x|=n$

$$p_H(x) = \sum_{q_0 \cdots q_n} \pi'(q_0, q_1) B'(x_1, (q_0, q_1)) A'((q_0, q_1), (q_1, q_2))$$

$$B'(x_2, (q_1, q_2)) \cdots A'((q_{n-2}, q_{n-1}), (q_{n-1}, q_n))$$

$$B'(x_n, (q_{n-1}, q_n))$$

$$= \sum_{q_0 \cdots q_n} \pi(q_0) A(q_0, q_1) B(q_0, x_1, q_1)$$

$$A(q_1, q_2) B(q_1, x_2, q_2)$$

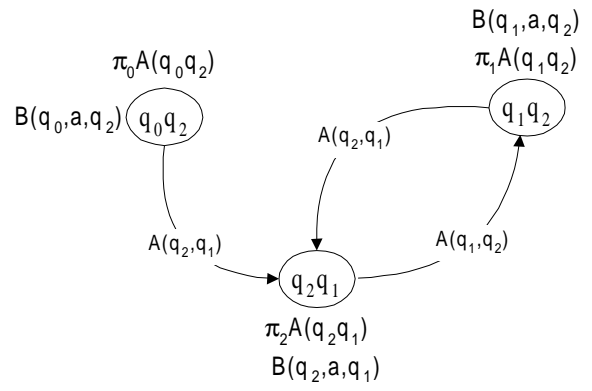
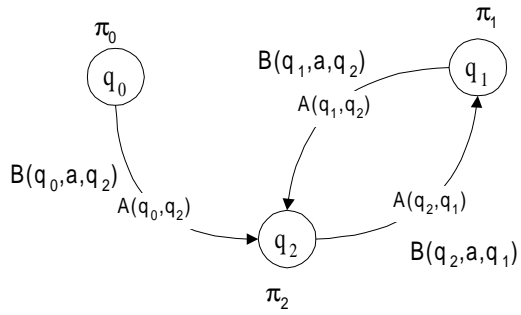
$$\cdots A(q_{n-1}, q_n) B(q_{n-1}, x_n, q_n)$$

$$= p_{HT}(x)$$



HMMT

HMM



⋮
 HMMTF - *HMM with observation probability distribution in the transitions and final state*
 ⋮

HMMTF -HMMTF com a restrição de um estado final “absorvente”

$$HTF = (Q, \Sigma, A, B, q_i, q_f)$$

$$\sum_{q'} A(q, q') = 1 \quad \forall_{q \in Q - \{q_f\}}$$

$$B: Q \times \Sigma \times Q \rightarrow [0,1]$$

$$\sum_{a \in \Sigma} B(q, a, q') = 1 \quad \forall_{q, q' \in Q}, \text{ se } A(q, q') \neq 0$$

$q_i \equiv$ estado inicial

$q_f \equiv$ estado final, no qual todas as transições são proibidas
 Todas as seqüências terminam no estado final

Probabilidade de observação da seqüência

$$x = x_1, x_2 \dots x_n, \quad x_i \in \Sigma$$

$$p_{HTF}(x) = \sum_{q_1 \dots q_{n-1}} A(q_i, q_1) B(q_i, x_1, q_1) \dots A(q_{n-1}, q_f) B(q_{n-1}, x_n, q_f)$$

HMMTF \Leftrightarrow SFSG

Proposição:

Dado um HMMTF $HTF = (Q, \Sigma, A, B, q_i, q_f)$ existe uma SFSG $G_e = (G, \mu)$, com $G = (N, \Sigma, R, S)$, tal que $\forall_{x \in \Sigma^*}, p_{G_e}(x) = p_{HTF}(x)$

$$\begin{aligned}
 N &= Q - \{q_f\} ; \quad S = q_i \\
 R &= \left\{ (q \rightarrow aq') \mid B(q, a, q') \neq 0 \text{ e } q' \neq q_f \right\} \\
 &\quad \cup \left\{ (q \rightarrow a) \mid B(q, a, q_f) \neq 0 \right\} \\
 \mu(q \rightarrow aq') &= A(q, q')B(q, a, q') \\
 \mu(q \rightarrow a) &= A(q, q_f)B(q, a, q_f)
 \end{aligned}$$

Proposição:

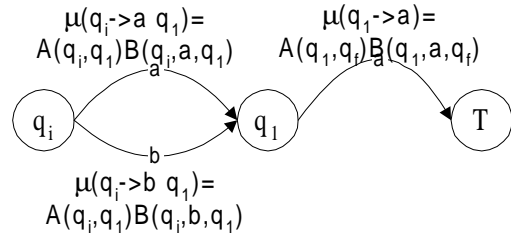
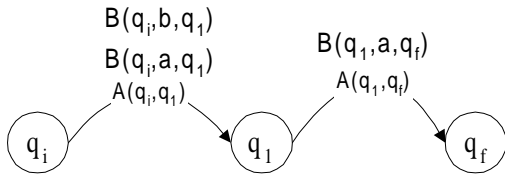
Dado uma SFSG, $G_e = (G, \mu)$, existe um HMMTF, $HTF = (Q, \Sigma, A, B, q_i, q_f)$, tal que $\forall_{x \in \Sigma^*}, p_{G_e}(x) = p_{HTF}(x)$

$$\begin{aligned}
 Q &= N \cup \{q_f\} & q_{f \notin N}, q_i &= S \\
 A(q, q') &= \sum_a \mu(q \rightarrow aq') & \forall_{q, q' \in N} & \\
 B(q, a, q') &= \mu(q \rightarrow aq') / A(q, q') & \forall_{q, q' \in N}, \forall_{a \in N} & \\
 A(q, q_f) &= \sum_a \mu(q \rightarrow a) & \forall_{q \in N} & \\
 B(q, a, q_f) &= \mu(q \rightarrow a) / A(q, q_f) & \forall_{q \in N}, \forall_{a \in N} & \\
 A(q, q') &= 0 & \text{se } (q \rightarrow aq') \notin R & \\
 A(q, q_f) &= 0 & \text{se } (q \rightarrow a) \notin R &
 \end{aligned}$$



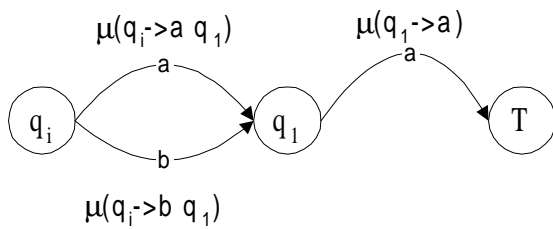
HMMT

SFSG

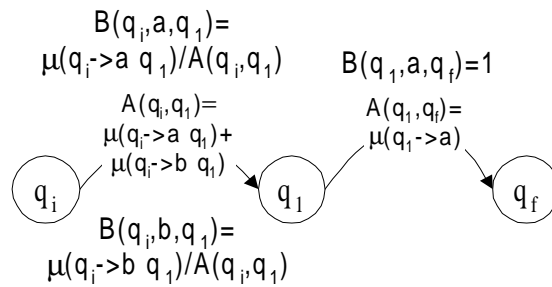




SFSG



HMMT





HMM

SFSG:

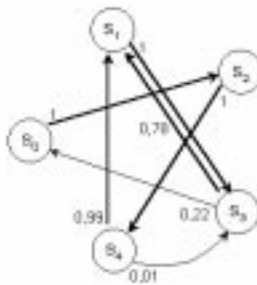
Prob. de observação de uma sequência:

$$p(x) = \sum_{q_1 \dots q_n} \prod_{t=1}^n A(q_{t-1}, q_t) B(q_t, x_t)$$

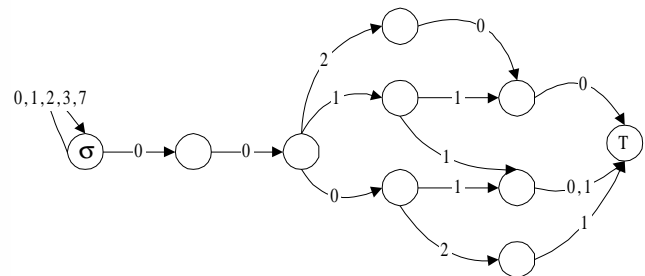
$$p(x) = \sum_{i=1}^D \prod_{t=1}^n \Pr(A_{t-1} \rightarrow x_t A_t)$$

Estrutura

- definida *a priori*



- inferida a partir do conjunto de treino



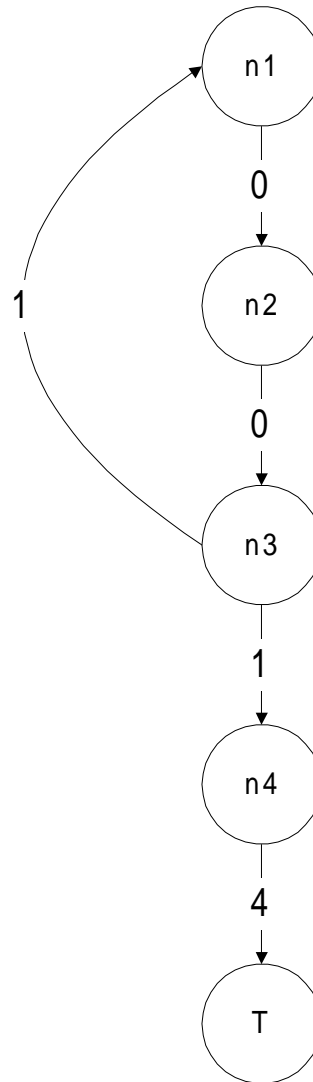
Estimação de Parâmetros

- Viterbi
- Baum-Welch (EM)

- Método da apresentação estocástica (equivalente a ML para gramáticas não ambíguas)

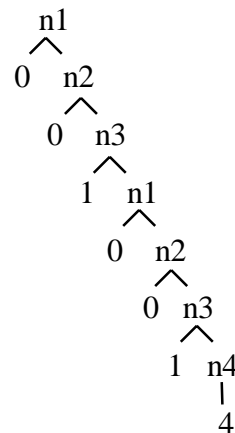
Gramáticas de Estados Finitos

- 0010010010014
- 0010010014
- $G=(N,\Sigma,R,S)$
- $N=\{n1,n2,n3,n4\}$
 $\Sigma=\{0,1,4\}$
 $S=n1$
- R: $n1 \rightarrow 0 n2$
 $n2 \rightarrow 0 n3$
 $n3 \rightarrow 1 n4 \mid 1 n1$
 $n4 \rightarrow 4$



- Derivação de uma sequência:

$n1$
 $0 n2$
 $00 n3$
 $001 n1$
 $0010 n2$
 $00100 n3$
 $001001 n4$
 0010014



SFSG - Gramáticas estocásticas de estados finitos

SFSG

$$G_e = (G, \mu)$$

$$G = (N, \Sigma, R, S) \text{ - gramática regular}$$

$$\mu : R \rightarrow [0,1]$$

$$\sum_{a \in \Sigma, B \in N \cup \{\epsilon\}} \mu(A \rightarrow aB) = 1$$

Derivação de x a partir de S de acordo com G é uma sequência de regras $D(x) = (r_1, r_2, \dots, r_{n(x)})$ que permite obter x a partir de S por sucessiva aplicação de regras em $D(x)$.

Probabilidade associada à derivação $D(x)$:

$$p(D(x)) = \mu(r_1)\mu(r_2)\dots\mu(r_{n(x)})$$

Probabilidade de geração da sequência

$$x = x_1, x_2 \dots x_n, \quad x_i \in \Sigma$$

$$p_{G_e}(x) = \begin{cases} 0 & \text{se não existe uma derivação para } x \\ \sum_{D(x)} p(D(x)) & \text{para todas as derivações } D(x) \end{cases}$$

Proposição: Dada uma SFSG $\sum_{x \in \Sigma^*} p_{Ge}(x) \leq 1$

$$P_n = 1 - \sum_{a_1 a_2 \dots a_n} \sum_{C_1 C_2 \dots C_n \in N} (\mu(S \rightarrow a_1 C_1) \dots \mu(C_{n-1} \rightarrow a_n C_n))$$

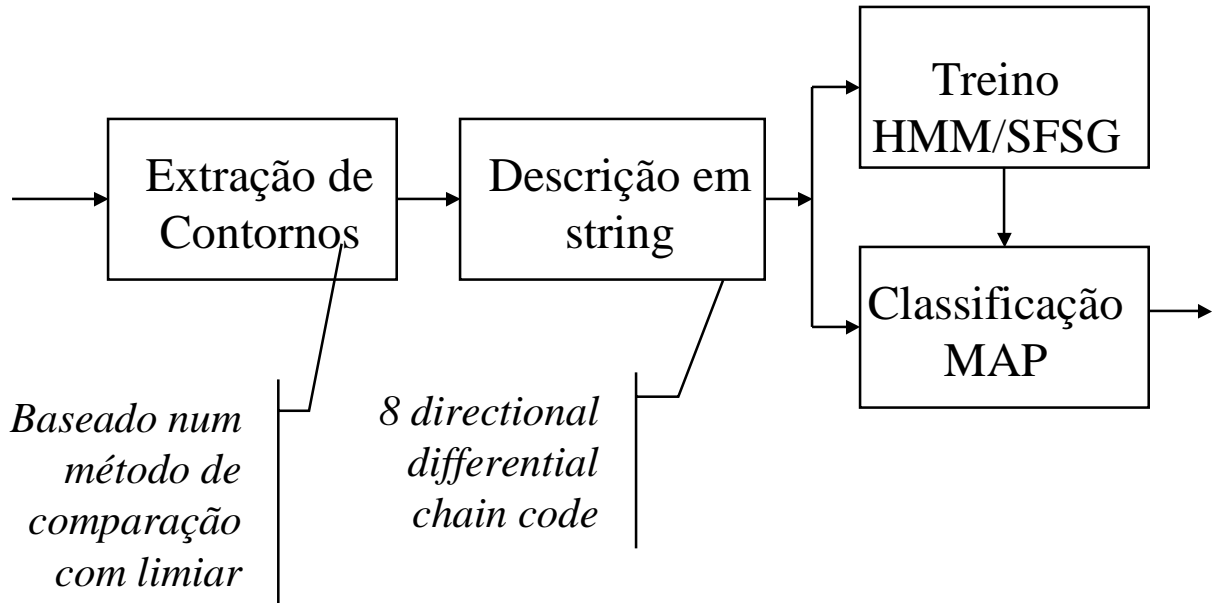
Quando $n \rightarrow \infty$, $P_n \rightarrow \sum_{x \in \Sigma^*} p_{Ge}(x)$

$$\therefore \sum_{x \in \Sigma^*} p_{Ge}(x) \leq 1$$

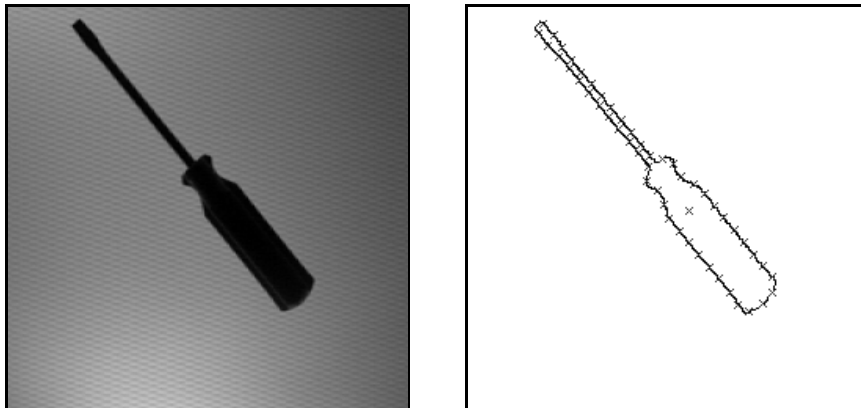
Dada uma SFSG, se $\sum_{x \in \Sigma^*} p_{Ge}(x) = 1$ a gramática diz - se consistente

Reconhecimento de Objectos

Metodologia



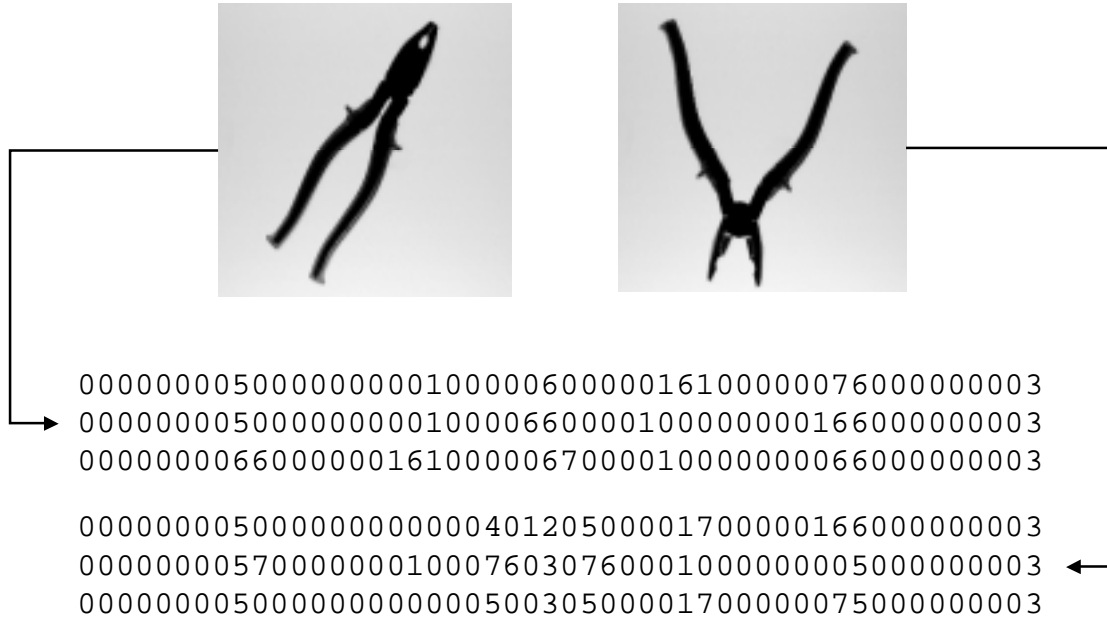
Extração de contornos



- O contorno do objecto é amostrado em 50 pontos equi-espaçados
- o ângulo entre segmentos consecutivos é quantificado em 8 níveis.

Base de Dados de Imagem

- 15 tipos de ferramentas
- 50 imagens por ferramenta, divididas em conjunto de treino e de teste
- incluem-se diferentes poses



Exemplos de Ferramentas



t4



t5



t6



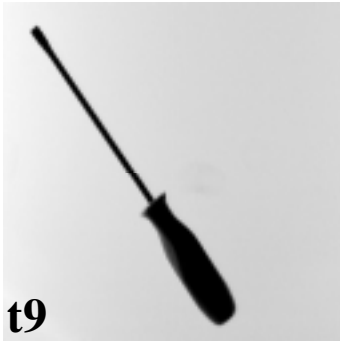
t7



t8



t9



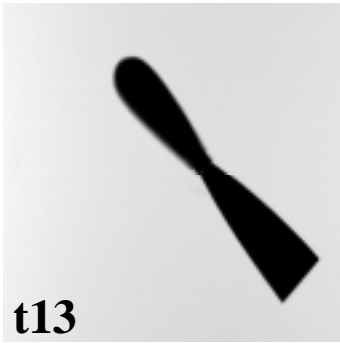
t11



t12



t13



t14



t15



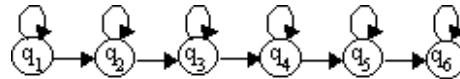
Aparato Experimental

Aprendizagem do Modelo

- Cada objecto é modelado por um HMM ou uma SFSG, treinado de acordo com:

HMM: Topologias:

- Totalmente ligada (10 & 20 estados)
- Esquerda-direita (20 & 50 estados)



Estimação de Parâmetros:

- Baum-Welch
- Viterbi

SFSG: Topologia:

- inferida a partir dos dados de treino usando o método das k -tails ($k=1, \dots, 10$)
- o número de estados depende da estrutura dos dados

Estimação de Parâmetros:

- Método da apresentação estocástica

Resultados

HMM: Os melhores resultados foram obtidos com o algoritmo de BW

	Ponto inicial fixo	Ponto inicial arbitrário
Totalmente ligada	99.7	99.5
Esquerda-direita	100	98.9

SFSG:

Test set	K=2			K=4			K=6		
	P_e	P_m	P_{ec}	P_e	P_m	P_{ec}	P_e	P_m	P_{ec}
Global	7.76	0	7.76	8.03	.3	7.76	9.14	5.3	3.88
t1	20	0	20	20	0	20	20	0	20
t2	0	0	0	0	0	0	0	0	0
t3	36	0	36	36	0	36	36	0	36
t4	0	0	0	0	0	0	0	0	0
t5	0	0	0	0	0	0	28	28	0
t6	0	0	0	0	0	0	14	14	0
t7	0	0	0	0	0	0	4	4	0
t8	33	0	33	33	0	33	0	0	0
t9	0	0	0	0	0	0	3	3	0
t10	0	0	0	0	0	0	0	0	0
t11	0	0	0	0	0	0	0	0	0
t12	0	0	0	0	0	0	7	7	0
t13	0	0	0	0	0	0	0	0	0
t14	0	0	0	0	0	0	0	0	0
t15	0	0	0	5	5	0	38	38	0

Pe: % erro; Pm: % não reconhecimento; Pec: % erro com prob-NN

Iter	Rec	t1	t2	t3	t4	t5-t7	t8	t9-t11	t12	t3-t15
1	92.24	k2	k2	k2	k2	k2	k2	k2	k2	k2
2	92.24	k3	k3	k3	k3	k2	k3	k2	k3	k2
3	92.24	k4	k4	k4	k4	k2	k4	k2	k4	k2
4	96.12	k5	k5	k5	k5	k2	k5	k2	k5	k2
5	96.12	k6	k6	k6	k6	k2	k5	k2	k5	k2
6	96.12	k7	k7	k7	k7	k2	k5	k2	k5	k2
7	96.12	k8	k8	k8	k8	k2	k5	k2	k5	k2
8	98.06*	k9	k9	k9	k9	k2	k5	k2	k5	k2
9	99.44*	k10	k10	k9	k10	k2	k5	k2	k5	k2
10	100*	k10	k10	k9	K8	k2	k5	k2	k5	k2

Rec: % global de reconhecimento

* Classificação *nearest-neighbor* probabilística

Conclusões

- A abordagem sintáctica permite uma automatização total dos processos de modelação e reconhecimento => as estruturas obtidas por SFSGs e HMMs são diferentes, as primeiras dependendo da complexidade estrutural dos dados.
- No respeitante à estimação de parâmetros o método de apresentação estocástica é semelhante ao algoritmo de Viterbi usado no treino dos HMMs.

Os resultados experimentais revelam:

- elevados níveis de reconhecimento por ambos os métodos
- HMMs são mais robustos no sentido em que possuem uma maior capacidade de generalização do que as SFSGs. Esta dificuldade é ultrapassada usando parsers correctores de erros (regra de decisão o vizinho mais próximo probabilístico) à custa de um maior custo computacional.
- As SFSGs conduzem geralmente a modelos de menor dimensão.