Capítulo 4

Um Método para Reconhecimento de Palavras Manuscritas do Extenso de Cheques Bancários Brasileiros

Os capítulos anteriores fornecem os subsídios necessários para que se possa apresentar os materiais e métodos desenvolvidos visando o reconhecimento de palavras manuscritas do extenso manuscrito em cheques bancários brasileiros. A revisão bibliográfica do Capítulo 2 apresenta os aspectos relevantes da escrita, do conhecimento prévio, da leitura e do reconhecimento. Além é claro, do estado da arte da área de Reconhecimento de Palavras Manuscritas. O Capítulo 3 descreve a formação da base de dado utilizada no presente trabalho.

Este capítulo está organizado de acordo com a seqüência de etapas necessárias ao desenvolvimento de um método de reconhecimento, como mostrado na Figura 1.2, sendo estas as seguintes:

- Pré-processamento: procedimentos preliminares, segmentação do extenso manuscrito em palavras isoladas, correção da inclinação vertical dos caracteres e detecção do corpo das palavras,
- Extração de primitivas: Perceptivas e Geométricas e
- Treinamento e validação dos modelos das palavras utilizando MEM e
- Reconhecimento das palavras do extenso manuscrito em cheques bancários utilizando MEM.

4.1. Pré-Processamento das Palavras Manuscritas

4.1.1. Procedimentos Preliminares

No Capítulo 2 relata-se a importância da definição dos parâmetros de aquisição das imagens. Portanto, para a aquisição das imagens dos extensos dos cheques de laboratórios foram adotados parâmetros com a finalidade de reduzir a possibilidade dos problemas gerados pela aquisição das mesmas. Após uma seqüência de testes com os 100 primeiros cheques de laboratórios da base de treinamento estes parâmetros puderam ser definidos. Diversos procedimentos de limiarização foram testados: Anisotropia, Entropia, Kittler, Otsu e ainda, a escolha manual do limiar. Diversos valores de brilho (70 a 90) e contraste (110 a 140) foram avaliados. Os parâmetros definidos foram:

- Brilho: 80 e
- Contraste: 128,
- Binarização: OTSU [OTSU, 1979].

Utiliza-se na aquisição das imagens um *scanner* FUJITSU M30096GX de alta resolução e grande autonomia para alimentação das folhas dos cheques de laboratório. Como o tipo de papel e de caneta utilizados no preenchimento dos cheques não apresenta grande diversidade, os parâmetros adotados obtiveram sucesso. Padronizou-se também a largura e altura das imagens coletadas, visando uma redução de espaço requerido para armazenamento e facilidade posterior no manuseio das mesmas.

Após aquisição e binarização, as imagens dos extensos foram submetidas a um processo de filtragem. Este processamento teve por objetivo melhorar o traçado das palavras (contorno externo). Adotou-se um processo de filtragem inicial das imagens proposto por [KOSKINEN et al., 1994] que consiste em uma filtragem denominada *soft.*

A filtragem *soft* caracteriza-se por ser uma técnica aplicável à imagens em níveis de cinza, sendo basicamente a operação de fechamento (dilatação seguida de erosão). A diferença principal está no fato de que um procedimento *soft* de fechamento em imagens em níveis de cinza permite trabalhar com qualquer um dos maiores valores dos vizinhos de um pixel qualquer. Desta forma, o resultado é mais suave do que em um procedimento clássico de fechamento. No caso das imagens do presente trabalho, o procedimento de filtragem *soft* foi aplicado automaticamente sobre as imagens binárias,

com uma iteração, para o primeiro maior valor de pixel vizinho. Como as imagens são binárias (preto ou branco), o maior valor será o do pixel branco.

As vantagens do procedimento proposto são as vantagens herdadas do procedimento de fechamento de imagens em níveis de cinza, de acordo com [FACON, 1995]:

- suaviza as fronteiras pelo exterior;
- preenche buracos;
- emenda partículas próximas e
- torna o conjunto mais regular.

A Figura 4.1 apresenta alguns exemplos de resultados obtidos com a aplicação da filtragem *soft*.

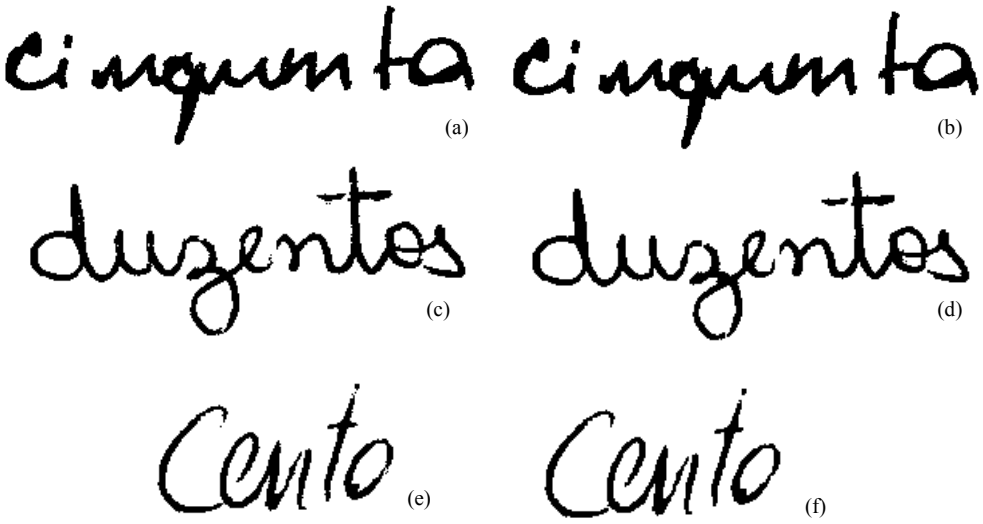

Figura 4.1: Exemplos de filtragem *soft*: a, c, e) imagens originais, b, d, f) imagens filtradas

4.1.2. Segmentação do Extenso Manuscrito em Palavras

A segmentação do extenso manuscrito em palavras possui dois objetivos, a saber:

- o de isolar as palavras para facilitar a realização da etapa de desenvolvimento dos MEM específicos de cada palavra do léxico,
- o de estabelecer uma estratégia de segmentação para o tratamento do extenso completo.

O presente trabalho refere-se ao primeiro objetivo, pois trata do reconhecimento das palavras isoladas e não do extenso como um todo.

Este processo de segmentação tem conhecimento das informações relativas ao tipo da escrita e dos símbolos iniciais e terminais, conforme apresentado no Capítulo 3. Porém inicialmente, o procedimento para segmentar o extenso em palavras não realiza nenhum tratamento específico para reconhecimento dos tipos de símbolos utilizados como elementos iniciais ou terminais do extenso propriamente dito.

A determinação do histograma de projeção vertical do extenso completo auxilia no processo de localização das diversas palavras que compõem o mesmo. Paralelamente, a determinação deste histograma são definidos dois tipos de espaçamento: o entre palavras e o intra-palavra.

O espaçamento entre palavras é por definição o espaço utilizado pelo escritor para separar duas palavras. E, o espaçamento intra-palavra é entendido como o espaço existente dentro de uma palavra, entre dois caracteres disjuntos. A Figura 4.2 apresenta a determinação do histograma de projeção vertical e exemplifica os dois tipos de espaçamentos definidos, isolando então, as diversas palavras contidas no extenso completo.

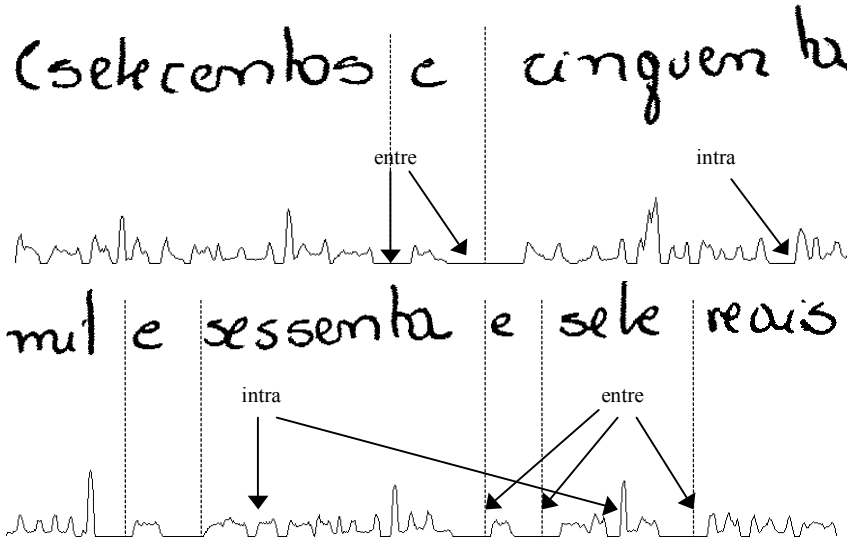

Figura 4.2: Exemplos de espaçamentos intra e entre palavras

Um primeiro procedimento de segmentação do extenso em palavras permitiu iniciar o estudo dos espaçamentos intra-palavra e entre as palavras. Esse estudo permite que este algoritmo de segmentação possa ser melhorado quando da sua aplicação no tratamento do extenso completo. O algoritmo utilizado para segmentar a base de dados

em palavras, evitando assim, o processo manual, tem por base o cálculo do tamanho médio de todos os espaçamentos horizontais existente em um extenso, obtidos pela análise do histograma de projeção vertical da imagem. Desta forma, os espaçamentos horizontais menores ou iguais a média são considerados, para o extenso em questão, como um espaçamento intra-palavra, não gerando segmentação. Por outro lado, os espaçamentos horizontais maiores que a média são considerados como espaçamento entre palavras, gerando segmentação.

O referido procedimento foi aplicado em toda a base de dados, gerando 11.948 palavras isoladas. Os erros existentes no processo foram os seguintes:

- segmentar uma palavra, devido a existência de um espaço intra palavra maior que a média dos espaçamentos existentes no extenso analisado e
- considerar uma palavra seguida de "e" ou "," como uma única palavra, devido a existência de um espaço entre palavras menor ou igual a média dos espaçamentos. Este erro ocorreu também, somente em dois casos, onde o escritor realizou a grafia das palavras muito próximas, por exemplo "cinqüenta e seis".

Outro ponto a ser considerado, são os símbolos iniciais e terminais utilizados pelos escritores. Na segmentação realizada não foram considerados como erro a segmentação de uma palavra associada a um símbolo inicial ou terminal. Todos esses casos foram resolvidos com mais uma passada pelo procedimento de segmentação.

Os dois tipos de erros ocorrem em 1076 palavras, representando 9% de erro no total das palavras segmentadas. Com isto, o resultado obtido mostra que o procedimento aplicado é eficiente, evitando o procedimento manual de segmentação do extenso em palavras.

Para o tratamento do extenso como um todo, este procedimento de segmentação do extenso em palavras isoladas deve ser modificado. Portanto, um estudo detalhado do espaçamentos existentes nas imagens dos extenso bancários deverá ser realizado. Com base nesses futuros estudos uma estratégia de segmentação do extenso em palavras poderá ser formulada. Essa estratégia refere-se a representação e modelagem dos espaçamentos, visando desenvolver um processo de segmentação robusto.

Um procedimento de segmentação robusto objetiva minimizar a segmentação excessiva ou *over-segmentation* e evitar segmentações que gerem palavras distintas falsamente unidas como uma única palavra ou *under-segmentation*.

A segmentação do extenso manuscrito em palavras pode parecer uma tarefa fácil para o ser humano, pois ao ler um texto consegue-se localizar palavra a palavra. No caso de dúvidas, basta ao ser humano aplicar o conhecimento prévio que possui e avaliar informações do contexto. Porém, o processo automático de localização de palavras pode se tornar uma grande dificuldade quando da identificação individual de palavras em um texto ou uma simples frase.

Segundo [PAQUET & LECOURTIER, 1993] assumir que os espaços entre palavras são sempre maiores que os espaços intra-palavras pode não ser válido para todos os casos. Fato este verdadeiro, uma vez que o procedimento aplicado resultou em 9% de erro. Ainda segundo os mesmos autores, este é o porque do processo de segmentação ser entendido como um provedor dos separadores das possíveis palavras, o que será confirmado ou não, dependendo da estratégia aplicada.

A estratégia mais comum é a chamada estratégia heurística que tem por base um processo de limiar global ou *global threshold*. Em resumo, os espaços com largura superior ao valor limite são considerados como separadores e os espaços com largura inferior ao valor limite são descartados.

Os estudos que podem ser realizados com os espaçamentos intra e entre palavras devem caracterizar-se por fornecer informações à modelagem dos espaçamentos através dos MEM, em lugar de aplicar-se uma estratégia heurística de segmentação do extenso em palavras. Desta forma, a melhor opção não é segmentar explícitamente, ou seja, o utilizar um valor fixo globalmente no extenso.

Outra estratégia é a recomendada por [PAQUET & LECOURTIER, 1993] na qual nenhuma decisão definitiva é tomada sobre os espaços em branco. E sim, os espaçamentos são considerados como fonte de diferentes formas de segmentação de um mesmo extenso. Neste caso, pode-se considerar a opção pela segmentação implícita, ou seja, que o extenso será tratado como uma única unidade com múltiplas opções de segmentação.

Uma proposta para este método é tratar o extenso manuscrito através das seguintes estratégias de segmentação:

- Segmentação Explícita: avaliar o desenvolvimento de MEM para modelar os espaçamentos intra e entre palavras, desde que os estudos realizados demonstrem a existência de dois conjuntos distintos de espaçamentos sobre a base de dados de treinamento;

- Estratégia Implícita: considerar todo e qualquer espaçamento como provável espaçamento entre palavras. Efetuar uma análise de espaçamentos ambíguos. Neste caso, serão considerados como espaçamentos ambíguos, da mesma forma que [PAQUET & LECOURTIER, 1993], aqueles que tanto podem ser modelados como espaçamento intra ou entre palavras.

De um modo geral, estas estratégias de segmentação fornecem ao método um mecanismo de adaptação aos diferentes aspectos de variabilidade da escrita cursiva humana.

4.1.3. Correção da Inclinação Vertical da Escrita

Com base na revisão bibliográfica, muitas técnicas podem ser aplicadas na correção da inclinação vertical dos caracteres, por exemplo, as técnicas apresentadas por [EL YACOUBI, 1996] e [GUILLEVIC, 1995] com base nos histogramas de projeção horizontal. Realizou-se a implementação de ambas as técnicas, iniciando pela técnica proposta em [EL YACOUBI, 1996], conforme descrição no Capítulo 2.

Em um segundo estudo, aplica-se o algoritmo dado por [GUILLEVIC, 1995]. Esta técnica tem como princípio a determinação dos histogramas de projeção horizontal das imagens, como na técnica proposta por [EL YACOUBI, 1996], para vários ângulos a partir do eixo vertical em passos de "delta_teta" graus. [GUILLEVIC, 1995] considera "delta_teta" igual a 5°.

Considerando inicialmente a premissa do valor de "delta_teta" da mesma forma que feito por [GUILLEVIC, 1995] percebe-se que para inclinações pequenas o algoritmo não gera alteração do ângulo de inclinação. Portanto, o algoritmo foi alterado para aceitar qualquer valor de "delta_teta", permitindo que testes fossem realizados.

Outra dificuldade do algoritmo de [GUILLEVIC, 1995] reside no caso de traços longos que podem influenciar a determinação do ângulo de inclinação uma vez que estes traços possuem um grande concentração de pixels. Portanto, o algoritmo proposto trabalha com *delta_teta* a partir de 1 grau e avalia inclinações entre 0 e ±90 graus. Desta forma algumas dificuldades encontradas pelo autor foram superadas. A Figura 4.3 apresenta alguns resultados obtidos com esta técnica.

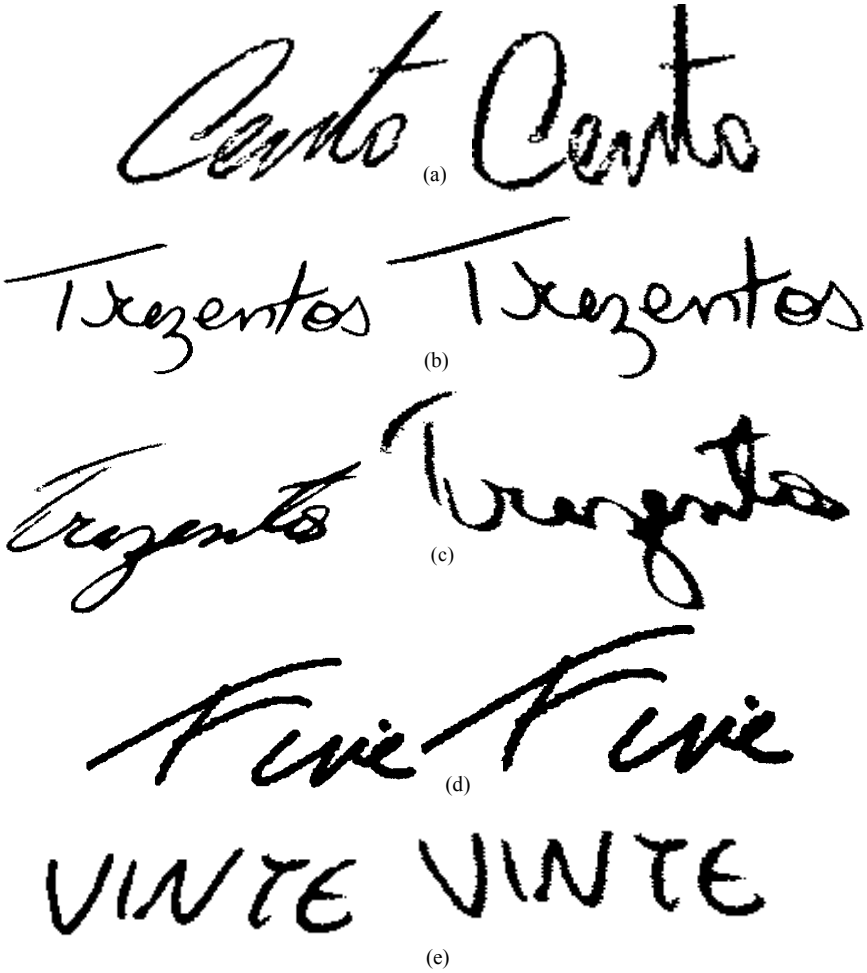

Figura 4.3: Exemplos de correção da inclinação vertical dos caracteres: a) correção de 40 graus, b) correção de 23 graus , c) correção de 58 graus , d) imagem extraída de [GUILLEVIC, 1995] com 7 graus de inclinação (ver resultado obtido por [GUILLEVIC, 1995] na Figura 2.22), e) correção de 29 graus

Ambas as técnicas fazem uma estimativa global da inclinação dos caracteres, sendo apresentadas como desvantagens, o tempo de processamento [EL YACOUBI, 1996] e [GUILLEVIC, 1995], e ainda, a influência na determinação do ângulo de inclinação para os casos de traços longos horizontais com inclinação diferente de 0 graus [GUILLEVIC, 1995].

Deste modo, opta-se por aplicar uma técnica simples, porém rápida, apresentada por [EL YACOUBI et al., 1999] baseada no contorno externo das palavras. Para todos os componentes da imagem extraí-se o contorno externo e a espessura média do traçado destes componentes. Percorre-se, então, o contorno externo com um passo igual a espessura e, para estes pontos do contorno, calcula-se o ângulo de inclinação médio, considerando-se todos os componentes da imagem analisada.

Após a correção da inclinação vertical dos caracteres aplica-se um procedimento de suavização da imagem da palavra de acordo com [STRATHY, 1993]. O objetivo é regularizar a continuidade do contorno das palavras, eliminando pequenos ruídos nas imagens. A Figura 4.4 mostra os resultados obtidos com a aplicação das etapas de correção da inclinação vertical das palavras e de suavização.

Figura 4.4: Pré-processamento das palavras: a)imagens originais e b)correção da inclinação dos caracteres e suavização do contorno externo

4.1.4. Detecção do Corpo da Palavra

A detecção do corpo da palavra permite estabelecer 3 zonas principais que são a base da relação entre a primitiva extraída e a sua localização na palavra. Portanto, as principais znas de uma palavra podem ser definidas como:

- **Z1 = zona ascendente:** compreendida entre o limite superior máximo (LSM) da imagem e o limite superior do corpo da palavra (LS);
- **Z2 = zona corpo da palavra:** compreendida entre o limite superior (LS) e inferior (LI) do corpo da palavra;
- **Z3 = zona descendente:** compreendida entre o limite inferior do corpo da palavra (LI) e o limite inferior mínimo da palavra (LIM).

A Figura 4.5 exemplifica as zonas de uma palavra manuscrita em estilo cursivo puro. Deve-se ressaltar que para palavras formadas por caracteres disjuntos contidos em retângulos pré-impressos (bastão ou caixa alta) [TAPPERT et al, 1990] os limites superior máximo e inferior mínimo da palavra coincidem com os limites superior e inferior do corpo da palavra, respectivamente.

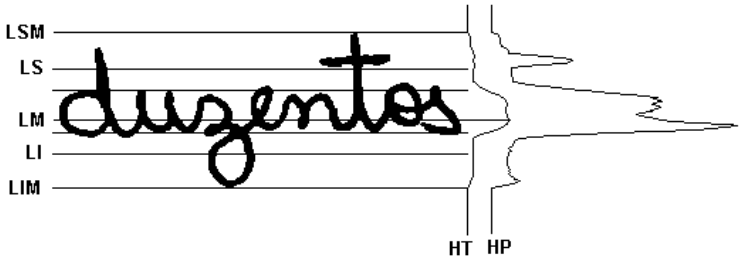

Figura 4.5: Zonas de uma palavra

A determinação das zonas ascendente, corpo da palavra e descendente é realizada através do histograma de projeção horizontal das transições (HT) branco-preto da palavra. Inicialmente, determina-se o pico máximo deste histograma, ou seja, a linha onde mais ocorrem transições branco-preto, denominada com Linha Média (LM). Aplica-se uma suavização no histograma para eliminar inflexões bruscas, seguindo a seguinte formulação:

$$P_x' = \frac{P_{x-2} + P_{x-1} + P_x + P_{x+1} + P_{x+2}}{5} \tag{4.1}$$

Com o histograma suavizado e a partir do pico (LM) buscam-se os limites LS e LI para 70% do valor do pico do histograma de transições. A determinação deste limiar heurístico baseou-se no estudo da diferença entre as coordenadas y dos picos do histograma de transição branco-preto (HT) e do histograma de densidade de pixels (HP), para 1.742 imagens das palavras de "um" até "dez" da base de treinamento.

Ao encontrar-se as coordenadas y dos pontos com 70% do valor de pico de HT, LS e LI são determinados acrescentando-se às coordenadas encontradas 50% da altura do corpo da palavra, conforme Figura 4.5. Isto deve-se ao fato de que nenhum procedimento de correção da linha de base está sendo aplicado, uma vez que, o preenchimento do valor extenso nos cheques conta com o auxílio de duas linhas de referência.

A Figura 4.6 exemplifica a determinação das diferentes zonas para palavras da base de dados de treinamento. Os exemplos sofreram a aplicação de todas as etapas de pré-processmaneto descritas anteriormente.

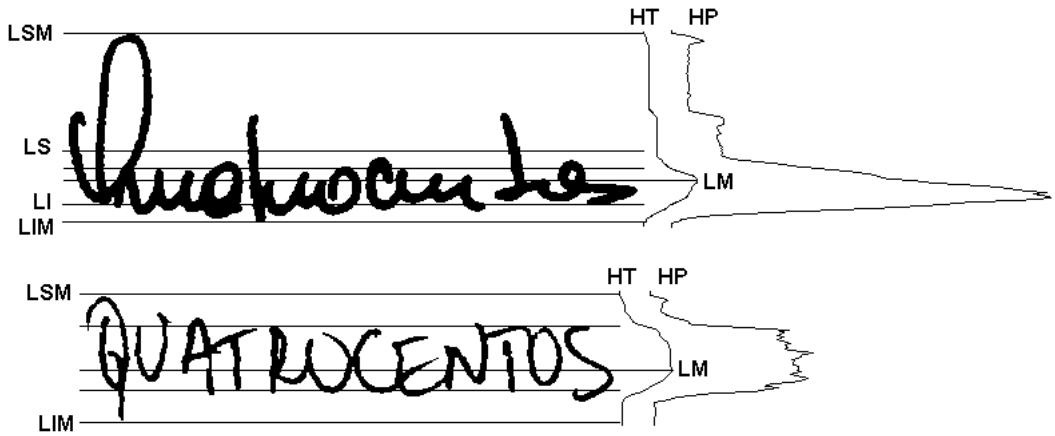

Figura 4.6: Determinação das zonas de palavras "quatrocentos"

O histograma de densidade de pixel não é utilizado para detecção do corpo das palavras uma vez que o mesmo gera distorções quando as palavras possuem uma concentração de pixels em local onde não se encontra o verdadeiro corpo da palavra, por exemplo: barra de corte da letra *t* e traçado irregular na escrita.

A seguir apresenta-se a extração de primitivas globais, compreendendo a seleção, o alfabeto de símbolos e as metodologias de extração de cada uma das primitivas.

4.2. Primitivas Perceptivas

A extração de primitivas caracteriza-se por uma das tarefas mais importantes no sucesso de métodos de reconhecimento de palavras manuscritas. É a partir das primitivas ou características das palavras escolhidas para serem extraídas que se obterá a robustez do método. Busca-se escolher um conjunto de primitivas invariantes, ou seja, que permanecem constantes para um mesmo estilo de escrita. E ainda, cada primitiva desse conjunto deverá ser independente, ou seja, não estar relacionada com as demais. Cada primitiva deve possuir identidade própria e expressar o significado de elementos que são característicos para o reconhecimento.

Existe uma variedade de primitivas que podem ser selecionadas para o processo de extração visando representar as formas a serem reconhecidas, como apresentado no Capítulo 2. O estudo destas primitivas permite compreender a complexidade e riqueza de informações contidas nas palavras manuscritas. Esta riqueza de informações, por sua vez, permite uma redução da complexidade do problema, quando explorada adequadamente.

No Capítulo 2, seção 2.3.5, apresenta-se uma discussão referente a seleção das primitivas a serem extraídas das palavras manuscritas. Leva-se em consideração os aspectos de escrita, leitura e conhecimento prévio, também apresentados no mesmo capítulo. Ressalta-se ainda os trabalhos de [MADHVANATH & GOVINDARAJU, 1998] e [SCHOMAKER & SEGERS, 1998] para o processo de seleção das primitivas globais.

Deve-se ter em mente que o enfoque do presente trabalho trata as palavras como uma unidade para o reconhecimento, ou seja, uma abordagem Global. Para trabalhar

com as palavras manuscritas extraíndo primitivas perceptivas foram selecionados 3 conjuntos de primitivas, a saber:

- Conjunto 01 – PP – Primitivas Perceptivas,
- Conjunto 02 – PPCC – Primitivas Perceptivas, Concavidades e Convexidades,
- Conjunto 03 – PPCCR – Primitivas Perceptivas, Concavidades e Convexidades Rotuladas.

A seguir descreve-se cada um dos conjuntos de primitivas selecionados e os respectivos processo de extração.

4.2.1. Conjunto 01 – Primitivas Perceptivas

Cabe a cada pesquisador buscar um conjunto de primitivas representativo, invariante, independente entre si, de tamanho adequado ao tempo de processamento e adequado ao método de reconhecimento a ser utilizado. Portanto, a seguinte pergunta deve ser respondida: ***Quais são as primitivas perceptivas relevantes no processo de reconhecimento de palavras manuscritas?*** O trabalho de [MADHVANATH & GOVINDARAJU, 1998] fornece a definição de primitivas perceptivas e o trabalho de [SCHOMAKER & SEGERS, 1998] fornece elementos que auxiliam na escolha das primitivas a serem extraídas.

Deste modo, o primeiro conjunto de primitivas proposto denomina-se Conjunto 01 - PP – primitivas percptivas, sendo composto por: ascendentes, descendentes e laços fechados nas 3 zonas de uma palavra, conforme Tabela 4.1. Estas primitivas geram informação sobre o tipo de primitiva, o tamanho relativo ao corpo da palavra e a informação posicional (em relação ao eixo x e em relação as zonas da palavra).

Tabela 4.1: Conjunto de Primitivas 01 - PP – Primitivas Perceptivas

Primitivas	**Conjunto 01 – PP**
Ascendente grande ou pequeno	T, t
Descendente grande ou pequeno	F, f
Laço grande ou pequeno no corpo da palavra	O, o
Laço superior ou inferior	j, l
Vazio	X

Após a detecção dos limites da palavra (LSM e LIM) e dos limites do corpo da palavra (LS e LI), as palavras são analisadas considerando-se as zonas: ascendente ou superior, corpo e descendente ou inferior.

A diferenciação da altura dos ascendentes, descendentes e laços no corpo em classes se faz necessária para auxiliar na representatividade das primitivas extraídas das palavras. O objetivo é diferenciar os ascendentes, descendentes e laços no corpo em duas classes distintas: pequeno e grande.

Para tanto, foi realizado um estudo da razão da altura dos ascendentes, descendentes e laços pela altura do corpo da palavra. Foram utilizadas as palavras de "um" a "dez" da base de treinamento. Realiza-se o cálculo da razão das alturas para dois histogramas:

- histograma de projeção horizontal de densidade dos pixels da imagem (HP) e
- histograma de projeção horizontal de transições branco/preto (HT para 50%, 60% e 70%).

Inicialmente, foram comparadas as quantidades de primitivas extraídas com base nas duas metodologias aplicadas (HP e HT) para posteriormente, analisar os resultados do estudo da razão da altura dos ascendentes, descendentes e laços pela altura do corpo da palavra. Com base nos resultados obtidos, tem-se que o HT-70% reflete as características das palavras em estudo melhor que o HP. Na seqüência, pode-se analisar o limiar de valores para a classificação dos ascendentes, descendentes e laços no corpo. Assim, as primitivas grandes são aquelas que apresentam um altura maior que 50% da altura do corpo da palavra, caso contrário, são ditas primitivas pequenas. Os laços superiores e inferiores não possuem indicação de tamanho, sendo válido o tamanho do ascendente ou descendente correspondente ao laço extraído. Este limiar, vem de encontro a Figura 2.9 relativa ao estilo de escrita cursiva para o processo de alfabetização no Brasil. Nesta Figura, observa-se que a proporção entre as letras maiúsculas e minúsculas, e portanto, ascendentes, descendentes e laços conferem à escrita as razões entre alturas denotadas pelo limiar de 50%.

Desde modo, dois procedimentos são então aplicados: Encontra-Laços e Encontra_Asc_Desc. O algoritmo **Encontra-Laços** tem a seguinte descrição:

Início

Pintura_pixels_brancos;

Centro_gravidade_laço_fechado (x_{CG},y_{CG});

x_laço = x_{CG};

// A coordenada x_{CG} é utilizada no momento da extração da seqüência de primitivas

y_laço = y_{CG};

// A coordenada y_{CG} é utilizada para cálculo da altura do laço em relação as zonas da palavra (ascendente, descendente e corpo)

Cálculo_Coordenadas_Máxima_Mínima_Altura_Laço(x_{CG}, $y_{máxima}$, $y_{mínima}$);

Cálculo_Diâmetro_Laço (x_{CG},y_{CG}, diâmetro);

Se (diâmetro > 2*espessura média do traço da palavra)

Então É laço;

Se (É laço)

Então Se (LS < $y_{máxima}$ ≤ LSM)

Então posição é superior;

Senão Se (LI ≤ $y_{máxima}$ e $y_{mínina}$ ≤ LS)

Então posição é corpo da palavra;

Senão Se (LI < $y_{mínima}$ ≤ LIM)

Então posição é superior;

Se (posição = corpo)

Então Tamanho_laço = $y_{máxima}$ - $y_{mínina}$;

Se (Tamanho_laço > 50% da altura do corpo)

Então tamanho é grande;

Senão tamanho é pequeno;

Fim_Encontra_Laços;

Por sua vez, o algoritmo **Encontra_Asc_Desc** está estruturado como a seguir:

Início

// Com base nos valores de LS e LI convenciona-se que toda parte da imagem que está localizada entre LSM e LS é ascendente e toda parte entre LI e LIM é descendente

y = LS;

Para ($x = x_{esquerda}$; $x \leq x_{direita}$; x=x+1) **faça** //(largura da imagem)

Encontra_Coordenadas_X_Asc(x, $x_{ascendente}$);

Fim_para;

y = LS;

Para ($x = x_{esquerda}$; $x \leq x_{direita}$; x=x+1) **faça**

Encontra_Coordenadas_X_Desc(x, $x_{descendente}$);

Fim_para;

Cálculo_Altura_Asc ($h_{ascendente}$);

Cálculo_Altura_Asc ($h_{ascendente}$);

Se ($h_{ascendente}$ > 50 % da altura do corpo da palavra)

Então tamanho_asc é grande;

Senão tamanho_asc é pequeno;

Se ($h_{descendente}$ > 50 % da altura do corpo da palavra)

Então tamanho_desc é grande;

Senão tamanho_desc é pequeno;

Fim_Encontra_Asc_Desc;

A Figura 4.7 exemplifica a extração de primitivas para dois exemplares, da base de dados de treinamento e de reconhecimento, respectivamente. Observa-se que o processo de extração de primitivas para a palavra “duzentos” permite representar as primitivas perceptivas. Porém, o mesmo não ocorre com o exemplo da palavra “quarenta”, percebendo-se, portanto, a influência do estilo da escrita no processo de extração de primitivas.

As primitivas perceptivas são extraídas sobre a imagem das palavras e um procedimento de pseudo-segmentação é aplicado para obter-se a seqüência de observações correspondente, como mostrado na Figura 4.7. A cada transição branco-preto sobre LM realiza-se uma observação e emite-se um conjunto de símbolos correspondentes as primitivas extraídas, caracterizando um grafema, conforme Tabela

4.1. São consideradas somente as transições branco-preto que não se encontram dentro de laços no corpo das palavras. A Figura 4.8 exemplifica um pseudo-segmento retirado da palavra “quarenta” mostrada na Figura 4.7. O caracter # representa um separador entre grafemas.

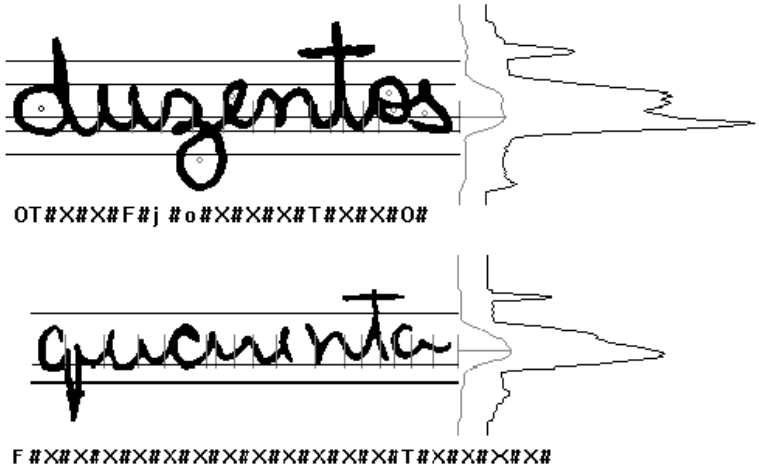

Figura 4.7: Extração de primitivas: conjunto 01 – PP – primitivas perceptivas

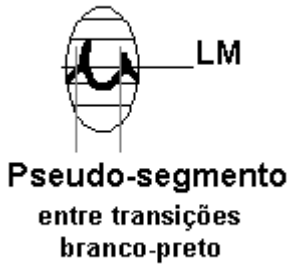

Figura 4.8: Pseudo-segmentação das palavras

Caso, nenhuma primitiva tenha sido extraída no segmento em análise, o símbolo X de vazio é emitido, de acordo com Tabela 4.1. O símbolo emitido independe da seqüência de ocorrência das primitivas no pseudo-segmento, uma vez que, a diferenciação entre as letras "b" , "d" e "p", "q" não é necessária. Pois, as letras "b" e "p" não fazem parte da formação das palavras que compõem o léxico.

4.2.2. Conjunto 02 – PPCC – Primitivas Perceptivas, Concavidades e Convexidades

Ao analisar-se os exemplos da Figura 4.7, palavra "quarenta", somente os ascendentes e descendentes são extraídos. Percebe-se que a representação da palavra torna-se pobre diante de tantas características existentes e perceptíveis aos olhos humanos.

Ao observar-se também, cuidadosamente, o vocabulário do extenso em cheques bancários brasileiros verifica-se a presença de um conjunto de palavras que não possui naturalmente primitivas perceptivas – ascendentes e descendentes, tais como: "um", "cinco", "seis", "nove", "cem", "reais".

Outro conjunto de palavras caracteriza-se por ser uma palavra curta, por exemplo: "dois", "três", "sete", "oito", "dez", "onze", "doze", "mil" e "real". Estas palavras possuem um grau de confusão natural, dependendo do estilo do escritor.

Deve-se ressaltar ainda, um terceiro conjunto de palavras que inclui as palavras manuscritas em letra tipo bastão ou de forma [TAPPERT et al., 1990], as quais não possuem as regiões de ascendentes e descendentes.

Assim a segunda pergunta deve ser respondida: ***Como representar palavras cursivas sem a presença de primitivas perceptivas ?***

[CÔTE, 1997] trata as palavras manuscritas como um conjunto de regularidades e singularidades. Classifica as singularidades como os ascendentes e os descendentes. E explora as concavidades e convexidades existentes no corpo das palavras para representação das regularidades, como mostrado na Figura 2.31 do Capítulo 2.

Assim, estuda-se a complementação do conjunto formado pelas primitivas perceptivas. Esta complementação tem por base a representação das concavidades e convexidades existentes no corpo das palavras [LECOLINET, 1990], [CHO et al., 1995], [LU & SHRIDHAR, 1996] e [CÔTE, 1997]. O segundo conjunto de primitivas é então denominado Conjunto 02 – PPCC - primitivas perceptivas, concavidades e convexidades, conforme Tabela 4.2.

As concavidades e convexidades são extraídas somente no corpo das palavras sobre o esqueleto das imagens e nos segmentos onde não occorem primitivas perceptivas. Desta forma, as concavidades e convexidades são primitivas complementares, auxiliando na representação das formas que não se classificam como

ascendentes, descendentes e laços fechados. Estas primitivas auxiliam na representação das curvaturas das letras e ligações entre letras, ou ainda, de laços abertos existentes no corpo das palavras.

Tabela 4.2: Conjunto de Primitivas 02 - PPCC - Primitivas Perceptivas, Concavidades e Convexidades

Primitivas	**Conjunto 02 - PPCC**
Ascendente grande ou pequeno	T, t
Descendente grande ou pequeno	F, f
Laço grande ou pequeno no corpo da palavra	O, o
Laço superior ou inferior	j, l
Côncavo	(
Convexo	)
Vazio	X

Os pontos convexos são determinados com o auxílio de uma família de 5 elementos estruturantes (matriz 7x3) e os pontos côncavos com uma família de 10 elementos estruturantes (matriz 9x4). Ambos os procedimentos de extração são aplicados sobre o esqueleto da imagem original pelo processo de *template matching*.

A definição dos elementos estruturantes teve por base a observação dos resultados de extração dos pontos côncavos e convexos para 1.742 imagens das palavras de "um" até "dez" da base de treinamento.

A Figura 4.9 mostra um exemplo da extração de primitivas considerando-se o Conjunto 02 – PPCC para as mesmas palavras da Figura 4.7. Observa-se que este conjunto de primitivas permite representar com mais informações as palavras que não possuem laços fechados, além do fato de caracterizar-se como adequado à extração de ligações e curvaturas naturais existentes em palavras manuscrita.

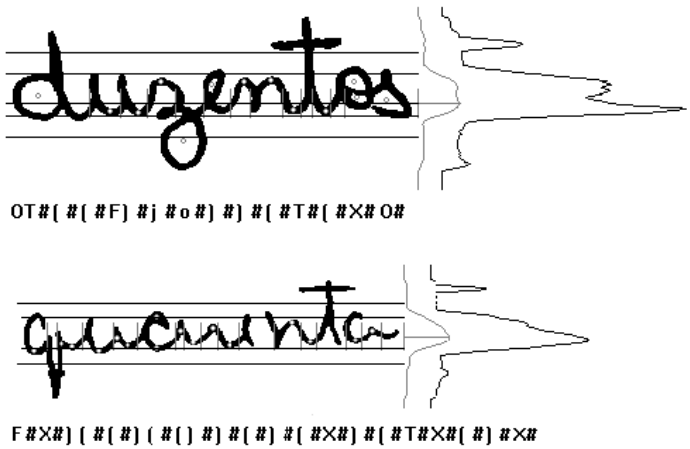

Figura 4.9: Extração de primitivas: conjunto 02 – PPCC

4.2.3. Conjunto 03 – PPCCR – Primitivas Perceptivas, Concavidades e Convexidades Rotuladas

Observa-se na base de treinamento que a proporção entre os grafemas extraídos como vazios, símbolo X, para o Conjunto de primitivas PP em relação aos grafemas com concavidades e/ou convexidades para o Conjunto PPCC é da ordem de 67.3%, fato que justifica a existência do Conjunto PPCC.

Por outro lado, a proporção dos grafemas com concavidades e/ou convexidades para o Conjunto PPCC em relação ao total de grafemas extraídos é da ordem de 50,5%. Portanto, percebe-se que a representação das concavidades e convexidades necessita de uma classificação mais detalhada para extração dos grafemas.

Deste modo, selecionou-se um terceiro conjunto de primitivas capaz de diferenciar grafemas extraídos de pseudo-segmentos compostos pelas letras “C”, “S”, “E” e “Z”, e ainda, “u”, “n”, “r” e “v”.

Este conjunto denominado de Conjunto 03 – PPCCR - primitivas perceptivas acrescido de concavidades e convexidades rotuladas, tem por base tratar as C*onvex Deficiencies* através de um procedimento de rotulação de pixels proposto por [PARKER, 1997], de acordo com a Tabela 4.3.

Tabela 4.3: Conjunto 03 – PPCCR - Primitivas Perceptivas, Concavidades e Convexidades Rotuladas

Primitivas	**Conjunto 03 - PPCCR**
Ascendente grande ou pequeno	T, t
Descendente grande ou pequeno	F, f
Laço grande ou pequeno no corpo da palavra	O, o
Laço superior ou inferior	j, l
Côncavo e Convexo (fechado)	a
Convexo (aberto embaixo)	n
Convexo (aberto a esquerda)	Z
Convexo (aberto a direita)	C
Côncavo (aberto em cima)	u
Ligação inferior	i
Ligação superior	r
C +)	S
C + (	E
Vazio	X

A rotulação é realizada a partir da análise de 4 direções (norte, sul, leste e oeste) para cada um dos pixels do fundo ou do *background* da área correspondente a parte

interna da concavidade e/ou convexidade extraída da imagem. Para [PARKER, 1997] existem 6 possibilidades de rotulação dos pixels, como descrito a seguir: a) Rótulo 0: todas as direções são fechadas, Rótulos 1 a 4: aberto em sul, aberto em oeste, aberto em norte, aberto em leste e Rótulo 5: quando o pixel não pode ser rotulado com nenhum dos casos anteriores.

Após a rotulação de todos os pixels da área em questão, os rótulos são contados e classifica-se a área com as seguintes opções: laço aberto, convexo do tipo "n", côncavo do tipo "u", côncavo do tipo "(" ou do tipo ")", convexo do tipo "C" ou do tipo "Z", convexo do tipo ligação superior "r" e côncavo do tipo ligação inferior "i". A Figura 4.10 exemplifica a extração deste conjunto de primitivas – PPCCR – para as mesmas palavras das Figuras 4.7 e 4.9. Observa-se que as palavras passam por um processo de enriquecimento de detalhes e representações ao comparar-se as Figuras 4.7, 4.9 e 4.10.

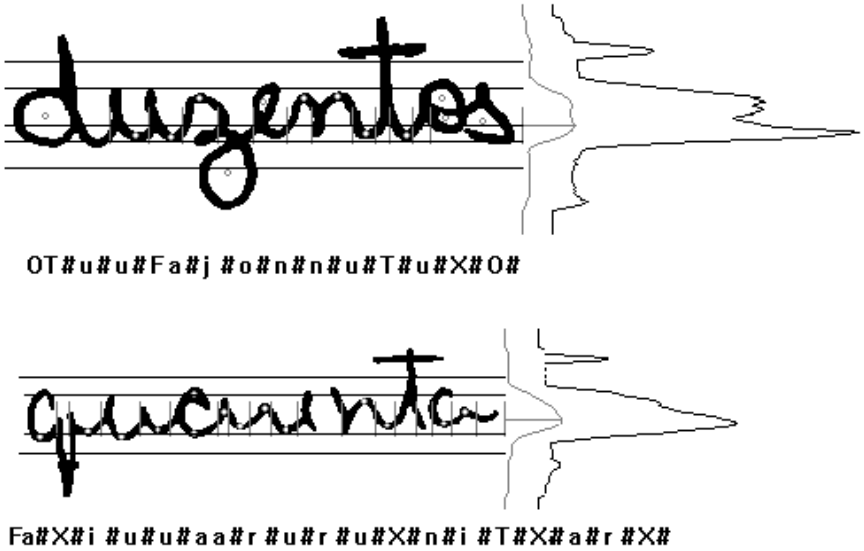

Figura 4.10: Extração de primitivas: conjunto 03 – PPCCR – primitivas perceptivas, concavidades e convexidades rotuladas

Assim, trabalha-se com a evolução de 3 conjuntos de primitivas globais buscando-se melhor representar as diferenças entre as palavras e as formas de escrita de uma mesma palavra. Ressalta-se que o presente trabalho não faz restrições ao tipo da escrita encontrada na base de dados utilizada.

Neste ponto, sabendo-se que para cada pseudo-segmento tem-se a extração de um grafema. Este grafema por sua vez representa a extração de nenhuma, uma ou mais

primitivas. Pergunta-se, então: ***Que associações entre as primitivas extraídas são relevantes?***

Diferentes conjuntos de primitivas e enfoques de validação podem ser encontrados na literatura. Estudos como os realizados por [BATTITI, 1994], [AVILA, 1996] e [GRANDIDIER et al., 2000] mostram a preocupação com a representatividade dos grafemas extraídos.

A seguir apresenta-se o estudo dos grafemas e a escolha dos alfabetos de símbolos que atendem as necessidades de representação das formas manuscritas no contexto de cheques bancários brasileiros.

4.2.4. Escolha de Alfabetos de Grafemas

A escolha de um alfabeto de grafemas necessita de dois passos importantes: o conhecimento de métodos de seleção dos grafemas e, principalmente, de uma definição de grafema. Esta definição é variável entre diferentes autores, pois depende diretamente dos métodos utilizados para descrever a palavra.

No presente trabalho, a noção de grafemas está diretamente ligada ao modo através do qual se realiza a extração de primitivas. O grafema corresponde ao traçado das entidades gráficas próximas da letra. Os pontos de pseudo-segmentação sobre a linha média (LM) das palavras, como apresentado anteriormente, definem a formação dos grafemas. Este enfoque assemelha-se ao dos pontos de ancoragem de [AVILA, 1996].

Define-se como conjunto de grafemas "bruto" o conjunto de todas as combinações de primitivas extraídas para os pseudo-segmentos e observadas sobre a base de dados de treinamento. Deve-se codificar o conjunto de grafemas "bruto" a fim de reduzir-se o número de símbolos possíveis para representar as palavras. Pois, para processo Markovianos de treinamento é interessante que o alfabeto de grafemas seja composto por um número limitado de símbolos representativos do problema. A operação de redução do conjunto de grafemas “bruto” permite estabelecer um alfabeto a ser codificado em símbolos, a partir da seqüência de observações extraídas das palavras, visando o processo de treinamento dos modelos através de MEM.

Apresenta-se a seguir o que é um grafema no contexto do presente trabalho e os métodos de estudo e definição dos alfabetos utilizados no trabalho. Os métodos aplicados são a Entropia de Shannon que permite uma medida da informação que cada

grafema carrega consigo e critérios de entropia para estudar as possíveis concatenações entre os grafemas.

4.2.4.1. O que é um Grafema?

Considera-se como um grafema uma entidade que pode corresponder a uma parte de uma letra, uma letra ou de letras conectadas. Anteriormente, apresentou-se que as primitivas são extraídas das palavras considerando-se as transições branco-preto sobre a linha média (LM), sendo que estas transições não podem estar localizadas dentro de laços fechados no corpo das palavras. Desde modo, uma palavra contém um número de grafemas correspondente ao número de transições branco-preto válidas sobre LM detectadas na palavra.

Supondo que os grafemas podem ser representados por um vetor de *n* elementos binários, onde cada elemento representa a ausência (0) ou a presença (1) de cada primitiva que compõe o Conjunto adotado, ter-se-ia conjuntos distintos de grafemas com as seguintes dimensões:

- Conjunto 01 – PP: $2^9 = 512$,
- Conjunto 02 – PPCC: $2^{11} = 2.048$,
- Conjunto 03 – PPCCR: $2^{18} = 262.144$.

Na realidade, foram observados na base de treinamento as seguintes configurações de grafemas, para um total de 58.432, extraídos de 7.146 amostras:

- Conjunto 01 – PP = 70 grafemas distintos,
- Conjunto 02 - PPCC = 74 grafemas distintos,
- Conjunto 03 - PPCCR = 94 grafemas distintos.

Diante deste fato, precisa-se melhorar a representação das palavras diante da grande variabilidade da escrita manuscrita, reduzindo-se assim o tamanho dos alfabetos de grafemas. Esta melhoria passa, então, pela análise dos conjuntos de grafemas "bruto" extraídos das palavras. Portanto, necessita-se de um processo que consiste, principalmente, em suprimir todas as configurações com baixa freqüência de aparição (por exemplo, configurações que ocorrem somente uma vez na base de treinamento).

Suprimir grafemas não significa eliminá-los, mas concatená-los com outros. Pois, o problema que se deseja evitar quando se trabalha com MEM pode ser respondido através da pergunta: *Como treinar modelos sobre grafemas que pouco são encontrados nas seqüências de observações?*

A concatenação de grafemas do conjunto "bruto" com baixa freqüência de aparição pode ser realizada ao definir-se um grafema que represente dois ou mais do conjunto "bruto" e supor-se que este novo grafema mais freqüente engloba os de baixa freqüência. Portanto, esta concatenação não deve reduzir a qualidade dos alfabetos definidos.

4.2.4.2. Escolha de uma Medida de Semelhança

A definição dos alfabetos parte do princípio que necessita-se de métodos que permitam estabelecer uma medida de semelhança entre as amostras, para assim poder concatenar os grafemas pouco frequentes e definir os alfabetos.

Quatro métodos podem ser utilizados na busca da semelhança entre os grafemas e definição de alfabetos, a saber:

- Distância de Hamming: permitir calcular as similaridades entre componentes presentes e ausentes;
- Distância de Hamming Ponderada: permite calcular a importância de cada uma das características dos grafemas representados por vetores de $\underline{n}$ elementos binários. Possui a desvantagem da atribuição empírica de pesos às diferentes primitivas consideradas;
- Alfabeto Hierárquico: introduz explicitamente uma hierarquia entre os grafemas de acordo com a freqüência das primitivas;
- Entropia de Shannon: este método estabelece uma medida da quantidade de informação aportada por cada um dos grafemas.

No presente trabalho os métodos Hierárquico e da Entropia de Shannon são adotados para estudo dos alfabetos a serem definidos. Esta escolha está ligada diretamente ao entendimento da freqüência de ocorrência dos grafemas e da quantidade de informação que cada um carrega consigo em relação aos demais.

A seguir apresenta-se a formulação da Entropia de Shannon e da Informação Mútua decorrente da aplicação no contexto dos grafemas estudados.

4.2.4.3. Entropia de Shannon e Informação Mútua

Seja uma variável aleatória discreta X, com $\{X=x_i\}$, $i \in \{0,...,N\}$ conjunto dos eventos de probabilidade de ocorrência P_i. A entropia de Shannon de X é uma medida

de informação que é definida como sendo [COVER & THOMAS, 1991] e [HUANG et al., 1990]:

$$H(P) = -\sum_{i=1}^{N} P_i \log_2(P_i) \tag{4.2}$$

Esta relação foi estabelecida por C. Shannon em 1948 e *H(P)* é, então, denominada de Entropia de Shannon. Esta medida de informação possui várias propriedades naturais, como as apresentadas a seguir:

- A entropia é máxima logo os eventos são equiprováveis;
- A entropia é mínima logo um dos eventos está correto. O sistema é perfeitamente conhecido e algum aporte de informação é possível;
- A entropia é uma função positiva, contínua e simétrica de seus argumentos;
- Para N eventos equiprováveis, a entropia é diretamente proporcional a N;
- A informação média adquirida entre duas etapas se ajusta por:

$$H(P_1 ... P_N) = H(Q_1, Q_2) + Q_1 H(\frac{P_1}{Q_1}, ..., \frac{P_M}{Q_1}) + Q_2 H(\frac{P_{M+1}}{Q_2}, ..., \frac{P_N}{Q_2}) \tag{4.3}$$

onde,

$$Q_1 = \sum_{i=1}^{M} P_i \quad \text{e} \quad Q_2 = \sum_{i=M+1}^{N} P_i \tag{4.4}$$

são as probabilidades de dois grupos de eventos *1...M* e *M+1...N*. A informação média do evento total é, portanto, a soma das informações médias associadas aos dois eventos da divisão ponderada pelas suas probabilidades de aparição e da informação de divisão.

O léxico para a aplicação dos extensos manuscritos de cheques bancários brasileiros é composto por 39 palavras diferentes. O valor de *H(P)* para o presente estudo foi calculado sobre a frequência relativa das palavras base da treinamento. Portanto, a entropia *H(P)* é:

$$H(P) = -P_0 \log_2 (P_0) - ..... P_{39} \log_2 (P_{39}) = 4{,}77 \text{ bits} \tag{4.5}$$

sendo a unidade de *H(P)* o bit. O bit é o valor de incerteza associado à seleção de um evento entre dois eventos equiprováveis (probabilidade = 0,5). O valor de entropia quantifica a variabilidade do problema. Este valor é a quantidade de informação mínima necessária a fim de poder resolver o problema de classificação de palavras. O valor da

entropia *H(P)* permite comparar o problema do presente trabalho com um problema de 27,28 classes equiprováveis, pois $\log_2 (27{,}28) = 4{,}77$.

Para definir a noção de Informação Mútua, deve-se introduzir duas variáveis aleatórias *X* e *Y* [COVER & THOMAS, 1991]. Nota-se $P(x_i)$ a probabilidade $P(X=x_i)$, $P(y_i)$ a probabilidade $P(Y=y_i)$, $P(x_i|y_i)$ e $P(y_i|x_i)$ as probabilidades condicionais $P(X=x_i|Y=y_i)$ e $P(Y=y_i|X=x_i)$.

A entropia da lei condicional é aquela que fornece a informação média aportada por *X* logo $Y= y_i$ é dado por:

$$H(X \mid Y = y_i) = -\sum_{i=1}^{N} P(x_i \mid y_i) \log_2 (P(x_i \mid y_i)) \tag{4.6}$$

A entropia condicional se deduz como a média obtida sobre Y:

$$H(X \mid Y) = -\sum_{j=1}^{M} P(y_j) H(X \mid Y = y_i)) \quad ou \tag{4.7}$$

$$H(X \mid Y) = \sum_{i=1}^{N} \sum_{j=1}^{M} P(y_j) P(x_i \mid y_i) \log_2 (P(x_i \mid y_i))$$

A entropia condicional quantifica a variabilidade do problema *X* quando *Y* é conhecido. A diferença entre *H(P)* e *H(X|Y)* define a quantidade de informação aportada por *X* sobre *Y* ou inversamente, pois entropia é simétrica. Esta informação é, então, denominada Informação Mútua entre *X* e *Y* e é definida por:

$$I_M(X,Y) = H(X) - H(X \mid Y) ou \tag{4.8}$$

$$I_M(X,Y) = H(Y) - H(Y \mid X)$$

que pode-se escrever sob a forma a seguir:

$$I_M(X,Y) = \sum_{i=1}^{N} \sum_{j=1}^{M} P(x_i, y_j) \log_2 \left(\frac{P(x_i, y_j)}{P(x_i) P(y_j)}\right) \tag{4.9}$$

A Equação (4.9) permite avaliar a quantidade de informação aportada individualmente por cada um dos grafemas.

O presente trabalho utiliza a Informação Mútua como uma medida para validar a redução do alfabeto de grafemas. Em um primeiro momento, utiliza-se para obter a informação mútua aportada por cada grafema “bruto” individualmente, identificando-se

os grafemas de maior representatividade das palavras permitindo uma análise global dos conjuntos de grafemas ditos “bruto”. Em um segundo momento, utiliza-se para realizar uma análise local, permitindo a validação dos grafemas concatenados através do método Hierárquico.

A aplicação da Informação Mútua para os grafemas toma como variáveis a palavras M e os grafemas G_k, sendo dada por [AVILA, 1996]:

$$I_M(M, G_K) = \sum_{i=1}^{N} \sum_{j=1}^{A} P(M = M_i, G_K = j) \log_2 \left(\frac{P(M = M_i, G_K = j)}{P(M = M_i) P(G_K = j)} \right) \tag{4.10}$$

onde, $j \in \{0,1,2,....,A\}$ representando o número de aparições do grafema k numa amostra de seqüências de observações da palavra M e $i \in \{0,1,2,....,N\}$ representando o número de palavras do léxico estudado.

4.2.4.4. Alfabetos Hierárquicos

Os alfabetos hierárquicos são assim denominados pois introduzem explicitamente uma hierarquia entre os grafemas segundo a freqüência das primitivas extraídas. Uma vez que, estes sejam organizados de maneira hierárquica permitem a conversão destes em símbolos, visando a obtenção das sequências de observações para entrada nos processos de treinamento e reconhecimento utilizando-se MEM.

Para formação de um alfabeto hierárquico deve-se considerar o conhecimento prévio da morfologia da forma a ser reconhecida, isto é, a palavra [AVILA, 1996]. Portanto, os conjuntos de primitivas devem auxiliar na descrição dos traçados das letras.

Desta forma, procura-se estabelecer símbolos que descrevem as letras do alfabeto possíveis de serem encontradas no léxico de palavras do extenso de cheques bancários brasileiros. Deve-se ressaltar que as seguintes letras não fazem parte da formação das palavras estudadas: b, f, g, p, x, k, w, y.

Assim, os alfabetos hierárquicos foram estabelecidos levando-se em consideração a freqüência de ocorrência dos grafemas “brutos” na base de dados de treinamento e as combinações mais apropriadas no que se refere a relação grafemas extraídos com as letras a serem representadas. Efetuaram-se, portanto, concatenações de dois ou mais grafemas dos conjuntos "brutos".

A Tabela 4.4 apresenta o alfabeto hierárquico para o Conjunto 01 de primitivas - PP. Este alfabeto é composto por 18 símbolos que descrevem os grafemas "brutos" mais freqüentes na base de dados de treinamento. [DOLFING, 1998] lembra que não é possível treinar um MEM com menos de 30 exemplares.

Tabela 4.4: Alfabeto Hierárquico – PP – Primitivas Perceptivas

Conversão Grafema-Símbolo	Descrição[1]
Se X **então** 01 **Senão** o **então** 02	Ausência Laço peq no corpo Laço grd no corpo Asc grd Asc peq Desc grd Desc peq Desc com laço inf Asc com laço sup Laço peq ou grd no corpo + Asc grd Laço peq ou grd no corpo + Asc peq Laço peq ou grd no corpo + Asc com laço sup Laço peq ou grd no corpo + Desc grd Laço peq ou grd no corpo + Desc peq Laço peq ou grd no corpo + Desc com laço inf Asc grd + Desc grd ou Asc grd + Desc peq ou Asc peq + Desc grd Asc peq + Desc peq Desc peq ou grd + Asc com laço sup Asc peq ou grd + Desc com laço inf Todas as demais opções

A Tabela 4.5 apresenta o alfabeto hierárquico para o Conjunto 02 de Primitivas – PPCC. Este alfabeto é formado por 22 símbolos distintos, incluindo-se os 18 símbolos do conjunto PP acrescido de 4 símbolos para representação das concavidades e convexidades no corpo das palavras.

Tabela 4.5: Alfabeto Hierárquico PPCC - Primitivas Perceptivas + Côncavo e Convexo

Conversão Grafema-Símbolo	Descrição
Senão (**então** 19	Côncavo no corpo (parte inferior)
Senão) **então** 20	Convexo no corpo (parte superior)
Senão (e) **então** 21	Côncavo e Convexo no corpo
Senão) e (**então** 22	Convexo e Côncavo no corpo
Senão **então** 18	Todas as demais opções

[1] asc (ascendente), desc (descendente), peq (pequeno), grd (grande), inf (inferior), sup (superior).

A Tabela 4.6 mostra o alfabeto hierárquico para o Conjunto 03 de Primitivas – PPCCR. Este alfabeto é formado por 29 símbolos diferentes, sendo os 20 primeiros símbolos provenientes do Conjunto de Primitivas PPCC, acrescido de 9 símbolos para representação das concavidades e convexidades rotuladas.

Tabela 4.6: Alfabeto Hierárquico PPCCR - Primitivas Perceptivas + Concavidades e Convexidades Rotuladas

Conversão Grafema-Símbolo	Descrição
Senão a **então** 21	Côncavo e Convexo (fechado)
Senão n **então** 22	Convexo (aberto embaixo)
Senão Z **então** 23	Convexo (aberto a esquerda)
Senão C **então** 24	Convexo (aberto a direita)
Senão u **então** 25	Côncavo (aberto em cima)
Senão i **então** 26	Ligação inf
Senão r **então** 27	Ligação sup
Senão S **então** 28	C +)
Senão E **então** 29	C + (

A composição destes alfabetos busca um equilíbrio na distribuição dos símbolos, oriundos dos conjuntos de grafemas “brutos”. Os alfabetos definidos proporcionam uma melhor distribuição da informação vinculada a estes símbolos, melhorando assim, o processo de treinamento, uma vez que reduziu-se também o número de símbolos de 70, 74 e 94 para 18, 22 e 29, respectivamente para os conjuntos PP, PPCC e PPCCR.

Fica a seguinte pergunta: *Como validar os alfabetos hierárquicos e garantir uma melhor distribuição da informação aportada pelos grafemas?* A seguir, apresenta-se o método de validação dos alfabetos hierárquicos através da Informação Mútua.

4.2.4.5. Validação dos Alfabetos Hierárquicos através da Informação Mútua

A validação dos alfabetos tem por base uma avaliação direta da quantidade de informação aportada pelos grafemas dos conjuntos "brutos" e dos concatenados pelo método hierárquico.

[GRANDIDIER et al., 2000] utiliza como indicador para a validação do conjunto de primitivas selecionado a perplexidade condicional baseada na noção de entropia. Outro exemplo, é o trabalho de [BATTITI, 1994] onde a Informação Mútua é aplicada para avaliar a informação aportada pelas primitivas e selecionar um sub-conjunto de primitivas representativas a ser utilizado como entrada de dados para um

classificador com Redes Neurais. [AVILA, 1996] aplica a Informação Mútua para melhorar o conjunto de primitivas utilizando um critério de concatenação de grafemas a partir de um conjunto “bruto”, visando o reconhecimento de extensos manuscritos de cheques franceses.

O método hierárquico permite que as concatenações entre grafemas corresponda a concatenação de mais de dois elementos. Assim, as concatenações podem ser validadas utilizando-se o seguinte critério descrito por [AVILA, 1996]:

$$\frac{I_M(M,G'')}{\max(I_M(M,G_1),I_M(M,G_2))} > \alpha \tag{4.11}$$

De outro modo, a relação entre a informação mútua do grafema resultante da concatenação $I_M(M,G'')$ e a maior informação mútua dos grafemas isolados deve ser superior ao limiar fixo α. O valor desta relação é superior a 1 se a concatenação vincula mais informação que o melhor dos grafemas isoladamente. Este valor é inferior e próximo de 1 se a concatenação vincula pouca informação.

Realiza-se a validação dos alfabetos hierárquicos aplicando-se a Equação (4.11) sobre os conjuntos dos "brutos" e codificados pelos símbolos das Tabelas 4.4 a 4.6, extraídos da base de treinamento.

A validação das concatenações para o conjunto de primitivas PPCCR está mostrada na Tabela 4.7. Pode-se tomar como exemplo o grafema hierárquico “*Ot*” (presença de um laço no corpo da palavra + um ascendente pequeno) corresponde a concatenação de quatro outros: “tO”, “Ot”, “ot” e “to”. O resultado “*Ot*” não faz diferença se o ascendente encontra-se antes ou depois do laço no corpo da palavra. Fato este, que vem de encontro a não necessidade de diferenciação das letras *d* e *b*, uma vez que a letra *b* não faz parte do alfabeto das palavras do léxico [FREITAS et al., 2001].

Os resultados mostram que os alfabetos hierárquicos são válidos, pois todas as concatenações realizadas apresentam $\alpha > 1$. Isto, significa uma melhoria na distribuição da informação aportada pelos grafemas com uma redução do tamanho dos alfabetos de símbolos.

A Informação Mútua mostra-se, portanto, um método rápido para uma inspeção direta sobre os grafemas dos conjuntos "brutos" e as concatenações realizadas, seja hierarquicamente ou por outro método. Permitindo, se necessário, que outros

procedimentos de concatenação possam ser aplicados, tais como os trabalhos de [BATTITI, 1994], [AVILA, 1996] e [GRANDIDIER et al., 2000].

Tabela 4.7: Validação das Concatenações para o Conjunto de Primitivas PPCCR

Grafemas "brutos"	max $I_M(M,G_k)$	Grafema Hierárquico	$I_M(M,G_k)$	Valor $\alpha > 1$
X	0,0096422	X	9,237108E-02	1,04385485
i	0,1575270	i	1,619584E+00	10,28131051
u	0,3182913	u	3,275907E+00	10,29216633
r	0,4338941	r	4,462884E+00	10,28565265
o	0,4754084	o	4,847420E+00	10,19632804
a,au,a),)a,aC	0,4829099	a	4,932612E+00	10,21435262
O	0,4846769	O	4,971346E+00	10,25703102
n,na,na	0,4583982	n	4,715003E+00	10,28582355
T	0,4100424	T	4,211878E+00	10,27181092
t	0,4089634	t	4,203431E+00	10,27825717
l	0,4107125	l	4,219515E+00	10,27364641
)	0,4084629	)	4,195261E+00	10,27084957
Z	0,4095092	Z	4,196184E+00	10,24686136
OT,TO,oT,To	0,4094520	OT	4,203069E+00	10,26510800
(	0,4086669	(	4,198055E+00	10,27255939
F	0,4097471	F	4,200450E+00	10,25132332
OF,FO,oF,Fo	0,4101486	OF	4,207193E+00	10,25772854
j	0,4088597	j	4,194596E+00	10,25925519
Ot,tO,ot,to	0,4102425	Ot	4,198101E+00	10,23321816
C	0,4095130	C	4,201715E+00	10,26027257
f	0,4090797	f	4,202851E+00	10,27391728
Ei,iE,Ea,aE,Er,rE,E	0,4104830	E	4,204452E+00	10,24269458
Ol,lO,ol,lo	0,4102881	Ol	4,207276E+00	10,25444316
jO,Oj,oj,jo	0,4104498	Oj	4,206434E+00	10,24835193
Ft,tF,ft,tf,fT,Tf,TF,FT	0,4104620	TF	4,209219E+00	10,25483236
of,fo,Of,fO	0,4102834	Of	4,209685E+00	10,26043218
lf,fl,Fl,lF	0,4103748	Fl	4,210608E+00	10,26039611
jt,tj,lj,jl,Tj,jT e grafemas com 3 primitivas*	0,4105242	Tj	4,213850E+00	10,26455931
S	0,4103597	S	4,281129E+00	10,43262533

* {TOf,fot,FoI,fto,loF,FtO,fOl,tFO,ljO,tOf,fol,ofT,Tfo,Oft,jot,FTO,flO,flo,tOj,tOF}

A seguir apresentam-se os experimentos de reconhecimento de palavras manuscritas dos extenso de cheques bancários brasileiros utilizando-se as seqüências de observações codificadas pelos alfabetos de símbolos definidos através de Modelos Escondidos de Markov - MEM.

4.2.5. Reconhecimento de Palavras Manuscritas do Extenso de Cheques Bancários Brasileiros

No Capítulo 2, fez-se uma revisão das atividades de escrita e leitura, além de uma associação destes processos com o conhecimento prévio. Foi apresentada também uma revisão dos aspectos de leitura e reconhecimento. Portanto, neste capítulo, foram

apresentadas as etapas que antecipam o reconhecimento de palavras manuscritas. É como se fosse necessário preparar as palavras para poder aprender sobre as mesmas e depois poder reconhecê-las.

Agora, pode-se apresentar a metodologia proposta para o reconhecimento das palavras isoladas do extenso manuscrito em cheques bancários brasileiros, bem como, detalhar os experimentos com MEM, suas dificuldades e soluções encontradas. Inicialmente, apresenta-se a fundamentação teórica dos MEM e os algoritmos utilizados nas etapas de treinamento, validação e reconhecimento, respectivamente.

A intenção é descrever as etapas de treinamento e validação dos 39 modelos estabelecidos para as palavras com uso de primitivas perceptivas, considerando-se um modelo para cada palavra do léxico. E, posteriormente, apresentar os resultados da etapa de reconhecimento obtidos com os 3 conjuntos de primitivas selecionados: PP, PPCC e PPCCR.

4.2.5.1. Fundamentação Teórica dos Modelos Escondidos de Markov

Os MEM resultam da associação de dois mecanismos ligados entre si: uma cadeia de Markov subjacente tendo um número finito de estados e um conjunto de funções de densidade de probabilidade de emissão de observações onde cada uma está associada com um estado. A cadeia de Markov serve de suporte a este conjunto de funções associadas a um estado.

Para a cadeia de Markov supõe-se que esta esteja em um estado qualquer e uma observação é emitida pela função de densidade de probabilidade condicionada pelo estado corrente. A cadeia de Markov troca então, de estado conforme sua matriz de probabilidade de transição. O observador vê somente os símbolos da observação e não pode diretamente observar os estados da cadeia de Markov, de onde o termo Modelos Escondidos de Markov – MEM, em *inglês Hidden Markov Model* – HMM.

A fundamentação teórica dos MEM está apresentada a seguir pelos seguintes tópicos:

a) Definição de MEM,
b) Tipos de Estruturas e
c) Problemas Básicos.

a) Definição de MEM

Para processos discretos, [HUANG et al., 1990] e [RABINER & JUANG, 1993] definem um MEM denotado por λ = (Q, V, A, B, π, T) pelos seguintes parâmetros:

T : comprimento da sequência de observações. Este comprimento pode ser variável,

Q = $\{q_1, q_2, ..., q_N\}$: o conjunto de estados do modelo. Q_t designará o estado da cadeia no instante t,

N = número de estado do modelo,

V = $\{v_1, v_2, ...v_m\}$: conjunto de símbolos de observações possíveis. O_t designará o símbolo da observação ao instante t,

M = número de símbolos da observação,

A = $\{a_{ij}\}_{1 \le i, j \le N}$: matriz de transição entre estados. O elemento a_{ij} corresponde a probabilidade que a cadeia de Markov transite do estado q_i para o estado q_j:

$$a_{ij} = pr(Q_{t=1} = q_j / Q_t = q_i) \qquad \forall i, j \in [1...N], \forall t \in [1, T-1] \tag{4.1}$$

com as seguintes restrições:

$$a_{ij} \ge 0 \quad \forall i, j \quad e \quad \sum_{j=1}^{N} a_{ij} = 1 \tag{4.2}$$

B = $\{b_{jk}\}_{1 \le j \le N;\ 1 \le k \le M}$: matriz de probabilidade de observação de símbolos condicionada pelos estados da cadeia. O elemento b_{jk} representa a probabilidade de observar o símbolo V_k quando o modelo se encontra no estado q_j:

$$b_j = pr(O_t = v_k / Q_t = q_j) \qquad 1 \le j \le N \qquad 1 \le k \le M \tag{4.3}$$

com as seguintes restrições:

$$b_{jk} \ge 0 \quad \forall j, k \quad e \quad \sum_{k=1}^{M} b_{jk} = 1 \quad \forall j \tag{4.4}$$

π = $\{\pi_i\}_{1 \le i \le N}$: conjunto de densidades de probabilidade inicial, π_i representa a probabilidade que o processo de Markov inicie no estado q_i:

$$\pi_i = pr(Q_1 = q_i) \qquad 1 \le i \le N \tag{4.5}$$

$$\pi_i \ge 0 \quad \forall i \quad e \quad \sum_{j=1}^{N} \pi_i = 1 \tag{4.6}$$

A avaliação desses processos de Markov é inteiramente determinada pela probabilidade inicial e pela probabilidade das transições sucessivas. O modelo de Markov assim definido é dito de primeira ordem.

b) Tipos de Estruturas de MEM

A determinação do MEM ótimo é uma tarefa que depende estritamente da estrutura do modelo e do número de estados escolhidos. De um modo geral, de acordo com [HUANG et al., 1990] e [RABINER & JUANG, 1993], existem 3 tipos de estruturas para os MEM:

- os modelos sem restrição, ditos ergóticos,
- os modelos seqüênciais e
- os modelos paralelos.

Os exemplos típicos desses modelo estão representados na Figura 4.11. Nos métodos sem restrição, Figura 4.11a, todas as transições possíveis entre os estado da cadeia são autorizadas. Isto é possível, se não se fizer restrição a nenhum dos valores a_{ij} igual nulo.

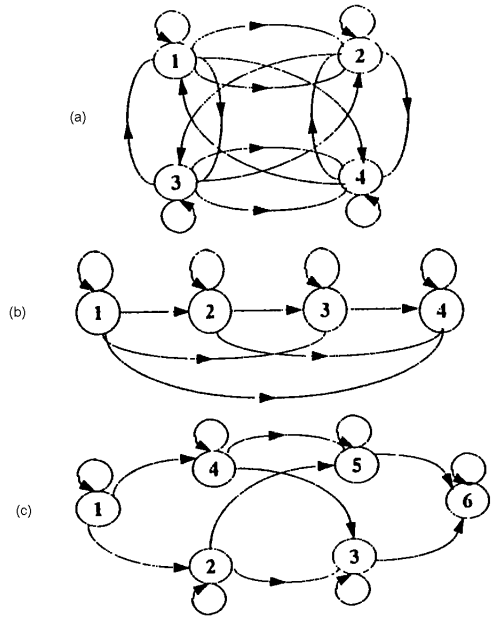

Figura 4.11: Tipos de estrutura dos modelos de Markov: a) modelo sem restrições, b) modelo sequencial, c) modelo paralelo [EL YACOUBI, 1996]

Os modelos seqüenciais e os modelos paralelos fazem parte dos modelos que seguem a *Topologia de Bakis*, ou ainda, demoninados de "left-right" ou "esquerda-

direita". Para esses modelos, a matriz de transição entre estados é triangular superior. Os modelos seqüenciais, Figura 4.11b, funcionam segundo uma evolução em série do modelo através de seus estados, mesmo que qualquer um desses estados possam ser saltados no curso do processo. Para os modelos paralelos, Figura 4.11c, muitos caminhos através da rede de Markov são permitidos, sabendo que cada um desses caminhos pode saltar um ou vários estados do modelo.

Cada uma das estruturas dos modelos da Figura 4.11 podem ser generalizadas por incluir um número de estados arbitrário. Entretanto, o número de parâmetros a estimar em um modelo de Markov é da ordem de N^2 (para a matriz A) mais NxM (para a matriz B). Por consequência, se N é muito grande, uma determinação coerente e precisa das matrizes A e B ótimas vem a ser muito difícil de realizar por uma base de treinamento de tamanho fixado.

Não existem meios teóricos pré-determinados de maneira a escolher o número N de estados necessários no modelo. Para [EL YACOUBI, 1996] a razão deste fato é que os estados nem sempre são fisicamente ligados aos fenômenos observáveis.

Além da variabilidade do tipo da estrutura do MEM, esses ainda podem ser classificados segundo o endereço onde são colocadas as observações, conforme [HUANG et al., 1990]:

- modelos de estados e
- modelos de arcos.

A Figura 4.12a mostra um exemplo de modelo de estados e a Figura 4.12b mostra um exemplo de modelo de arcos. No caso dos modelos de arcos, as duas probabilidades, a_{ij} e $b_j(k)$ consideradas no modelo de estados são substituídas por $pr_{ij}(k)$ que corresponde a probabilidade de transição do estado q_i ao estado q_j emitindo o símbolo da observação v_k.

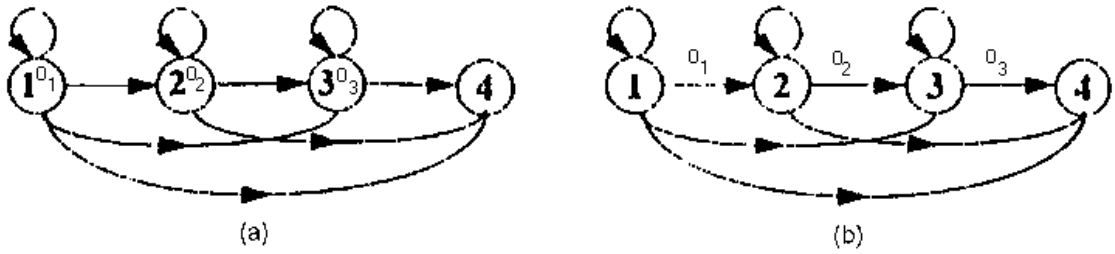

Figura 4.12: Modelo de estados e modelos de arcos: a) modelo de estados e b) modelo de arcos [EL YACOUBI, 1996]

c) Problemas Básicos

Os MEM são aplicados para responder a três problemas básicos [RABINER & JUANG, 1993]:

- Dada uma seqüência de observação e o modelo λ, computar eficientemente a probabilidade da seqüência de observações. Este é o problema da avaliação da probabilidade de observação,
- Dado um conjunto de seqüências de observações relativas as palavras da base de dados de treinamento, produzir um ou mais MEM para cada palavra. Este é o problema de treinamento dos parâmetros dos modelos.
- Dado um conjunto de MEM e uma seqüência de observações correspondente a uma palavra desconhecida, dar uma interpretação a esta seqüência atribuindo esta palavra a um dos modelos (ou classes) do sistema. Este é o problema de reconhecimento ou de classificação.

Existem duas maneiras principais de verificar a evolução da probabilidade de observação: por evolução direta e por evolução dos algoritmos de *Forward-Backward* [RABINER & JUANG, 1993]. Estes algoritmos serão apresentados a seguir.

O problema de treinamento é mais complexo. Felizmente, existem algoritmos potentes que garantem a convergência na direção dos modelos ótimos.

O problema de reconhecimento é resolvido calculando-se a probabilidade que cada modelo tem, produzindo-se à seqüência de observações correspondente a palavra desconhecida e atribuindo-se à esta palavra o modelo que fornece a maior verossimilhança. Esta é a classificação por máxima verossimilhança.

A seguir descreve-se o conjunto de condições iniciais para o estabelecimento dos modelos das palavras do léxico dos cheques bancários brasileiros, o treinamento, a validação e os experimentos de reconhecimento propriamente ditos.

2.4.5.2. Condições Iniciais

Considerando-se que a coleta da base de dados de laboratório não está terminada e o total de 11.948 palavras isoladas, optou-se pela divisão do conjunto de seqüências em três sub-conjuntos: 60% para treinamento, 20% para validação e 20% para reconhecimento.

A Tabela 4.8 apresenta o total de seqüências de observações obtidas para cada uma das palavras e a dimensão dos sub-conjuntos de treinamento, validação e reconhecimento, respectivamente.

Tabela 4.8: Conjuntos de Seqüências de Observações ([1] PC = palavras-chaves)

	Palavras	Total	T	V	R
UNIDADE	Um	332	199	67	66
	Dois	350	209	71	70
	Três	336	201	68	67
	Quatro	322	193	65	64
	Cinco	332	199	67	66
	Seis	302	181	61	60
	Sete	267	160	54	53
	Oito	305	182	62	61
	Nove	318	190	65	63
	%	**23,97**	**23,98**	**23,89**	**24,00**
DEZENA	Dez	47	28	10	9
	Onze	50	29	11	10
	Doze	29	17	7	5
	Treze	37	22	8	7
	Quatorze	35	20	8	7
	Quinze	45	26	10	9
	Dezesseis	30	17	7	6
	Dezessete	36	21	8	7
	Dezoito	34	20	8	6
	Dezenove	45	26	10	9
	%	**3,25**	**3,16**	**3,58**	**3,16**
ENTA	Vinte	382	229	77	76
	Trinta	372	223	75	74
	Quarenta	338	202	69	67
	Cinqüenta	376	225	76	75
	Sessenta	356	213	72	71
	Setenta	370	221	75	74
	Oitenta	363	217	74	72
	Noventa	316	189	64	63
	%	**24,05**	**24,06**	**23,98**	**24,08**
ENTOS	Cem	7	4	2	1
	Cento	269	161	55	53
	Duzentos	249	149	51	49
	Trezentos	236	141	48	47
	Quatrocentos	218	130	45	43
	Quinhentos	240	143	49	48
	Seiscentos	219	131	45	43
	Setecentos	248	148	51	49
	Oitocentos	252	151	51	50
	Novecentos	236	141	48	47
	%	**18,25**	**18,23**	**18,42**	**18,15**

Tabela 4.8: Conjuntos de Seqüências de Observações ([1] PC = palavras-chaves) - Continuação

	Palavras	Total	T	V	R
P C[1]	Mil	1.286	771	258	257
	Reais/Real	1.262	757	253	252
	Centavos/Centavo	1.094	656	220	218
	%	**30,48**	**30,56**	**30,12**	**30,61**
	Total	**11.948**	**7.146**	**2.427**	**2.375**

A seguir descreve-se cada um dos algoritmos e critérios adotados nas etapas de treinamento, validação e reconhecimento das palavras.

2.4.5.3. O Treinamento: Algoritmo *Baum-Welch*

Na fase de treinamento são necessárias técnicas que permitem captar a informação estatística sobre a variabilidade das formas que representam uma mesma palavra. O interesse dos MEM resulta da existência de algoritmos eficazes que permitem a convergência dos parâmetros dos modelos. Partindo de um valor inicial dos parâmetros, esses algoritmos interativos alinham os modelos sobre observações conhecidas (base de dados de treinamento) e calculam as estatísticas quando do momento dessas produções.

A partir dessas estatísticas, determina-se novos valores dos parâmetros que melhoram a probabilidade total da emissão das observações da base de dados de treinamento. Esta base de dados deve possuir uma dimensão que possibilite que cada fenômeno (forma) particular de variabilidade seja suficientemente representado.

O treinamento Markoviano está ligado a um contexto Bayesiano onde procura-se o modelo λ que melhor considera os dados, isto quer dizer é o mais provável em função dos dados. Procura-se portanto, λ tal que:

$$\hat{\lambda} = \arg\max \left(\Pr(\lambda / O_i, i = 1, \cdots, t)\right) \tag{4.7}$$

sendo $\hat{\lambda}$ o modelo estimado e O_i os dados do treinamento.

Pela regra de Bayes:

$$\Pr(\lambda / O_i, i = 1, \cdots, t) = \frac{\Pr(O_i / \lambda) \times \Pr(\lambda)}{\Pr(O_i)} \tag{4.8}$$

a otimização desta equação é realizada sob os modelos λ, onde obtém-se, portanto:

$$\hat{\lambda} = \arg\max \left(\Pr(O_i / \lambda) \times \Pr(\lambda)\right) \tag{4.9}$$

pois, os dados são por definição independentes:

$$\hat{\lambda} = \arg\max\left(\Pr(\lambda) \times \prod_i \Pr(O_i / \lambda)\right) \tag{4.10}$$

na qual *pr(λ)* é um meio pelo qual se faz preferir um modelo a outros. Em geral, a escolha é realizada utilizando-se,

$$\hat{\lambda} = \arg\max \prod_i \Pr(O_i / \lambda) \tag{4.11}$$

o que resulta em fazer um treinamento por Máxima Verossimilhança.

O treinamento supõe que todos os modelos são a priori equiprováveis. É difícil, então, estabelecer um critério para estimar a probabilidade *a prior*i do modelo.

Dois algoritmos podem ser utilizados na etapa de treinamento dos modelos: Baum-Welch e Viterbi. [EL YACOUBI, 1996] afirma que no processo de treinamento o algoritmo de Viterbi não leva em conta os eventos correspondentes ao alinhamento ótimo entre as formas observáveis e o modelo. O algoritmo de Baum-Welch leva em consideração todos os alinhamentos possíveis e maximiza a verdadeira probabilidade de emissão das observações. Por conseqüência, a desigualdade $P^V \leq P^{BW}$ é verdadeira, sendo a igualdade satisfeita somente quando a seqüência de estados que produz as observações é única.

A seguir descreve-se o algoritmo de Baum-Welch utilizado no escopo do treinamento dos modelos, por máxima verossimilhança, para o presente trabalho. A aplicação deste algoritmo está ligada ao fato do mesmo considerar todos os alinhamentos possíveis e maximizar a verdadeira probabilidade de emissão das observações. Isto é importante para o treinamento de formas que envolvem tantos fatores de variabilidade como os encontrados nas palavras manuscritas. Não apresenta-se o algoritmo de Viterbi, uma vez que não se aplica o mesmo, podendo ser encontrado em [HUANG et al., 1990] e [RABINER & JUANG, 1993].

O Algoritmo de Baum-Welch no Treinamento

Num primeiro instante, supõem-se de que somente uma seqüência de observações será processada. Isto por duas razões, por um lado para simplificar o problema, e por outro lado porque a extensão para múltiplas seqüências de observações é direta.

A idéia principal sobre a qual está baseada o algoritmo de Baum-Welch é que sendo dada uma estimação qualquer λ do modelo, e uma seqüência de observações O, as melhores estimativas dos parâmetros do novo modelo λ^{NM} são dadas por [HUANG et al., 1990] e [RABINER & JUANG, 1993]:

$$a_{ij}^{NM} = \frac{P(\textit{transição do estado } q_i \textit{ ao estado } q_j / O, \lambda)}{P(\textit{transição do estado } q_i \textit{ a um estado qualquer} / O, \lambda)} \tag{4.12}$$

$$b_{jk}^{NM} = \frac{P(\textit{emissão do símbolo } v_k \textit{ ao estado } q_j / O, \lambda)}{P(\textit{emissão de um símbolo qualquer a paritr do estado } q_j / O, \lambda)} \tag{4.13}$$

$$\pi_i^{NM} = P(\textit{seqüência de observação começando pelo estado } q_i / O, \lambda) \tag{4.14}$$

O interesse do algoritmo de Baum-Welch reside no fato de que as fórmulas de reestimação de A,B e π acima, garantem o acréscimo de *P(O/λ)* até que um ponto de convergência seja atingido, ponto para o qual os parâmetros não mudam mais. Este procedimento de controle do treinamento encontra-se apresentado através do Processo de *Cross-Validation* descrito posteriormente.

O algoritmo de Baum-Welch é composto por dois algoritmos: *Forward* e *Backward* que permitem calcular as probabilidades exigidas pelas Equações (4.12) a (4.14). A seguir apresenta-se esses dois algoritmos para posteriormente retornar ao cálculo das equações mencionadas, que por sua vez permitem obter os parâmetros do modelo em treinamento.

a) O Algoritmo de *Forward*

O algoritmo de Baum-Welch necessita do cálculo da probabilidade chamada *Forward*, que são as probabilidades conjuntas de emitir a seqüência de observação parcial $O_1, O_2..., O_t$ e de ocupar o estado q_i no instante *t*. Designa-se as probabilidades *Forward* por $\alpha_t(i)$, então:

$$\alpha_t(i) = P(O_1 O_2 \cdots O_t, q_t = i / \lambda) \tag{4.15}$$

A probabilidade que deseja-se calcular *P(O/λ)* é obtida sob a seguinte forma (a notação desta probabilidade P^{BW} fará referência ao algoritmo de Baum-Welch):

$$P^{BW} = \sum_{i=1}^{N} \alpha_T(i) \tag{4.16}$$

porque $\alpha_t(i)$ é a probabilidade de emitir O em conclusão ao estado q_i. Os α_t podem ser calculados recursivamente pelo algoritmo a seguir:

Início

1) INICIALIZAÇÃO

$$\alpha_1(i) = \pi_i b_i(O_1) \qquad 1 \le i \le N \tag{4.17}$$

2) INDUÇÃO

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij} \right] b_j(O_{t+1}) \quad 1 \le t \le T-1 \quad e \quad 1 \le j \le N \tag{4.18}$$

3) TÉRMINO

$$P(O/\lambda) = \sum_{i=1}^{N} \alpha_T(i) \tag{4.19}$$

Fim_Forward;

b) Algoritmo de *Backward*

Para calcular essas quantidades, as probabilidades *Forward* $\alpha_t(i)$ são completadas pelas probabilidades *Backward*, de acordo com o algoritmo seguinte:

Início

1) INICIALIZAÇÃO

$$\beta_T(i) = 1 \qquad 1 \le i \le N \tag{4.20}$$

2) INDUÇÃO

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j) \qquad T-1 \ge t \ge 1 \quad 1 \le i \le N \tag{4.21}$$

Fim_Backward;

Assim, o numerador da equação (4.12) é dado por:

$$P(\text{transição do estado } q_i \text{ ao estado } q_j, O/\lambda) = \sum_{t=1}^{T} \alpha_t(i) \beta_{t+1}(j) b_j(O_{t+1}) a_{ij} \tag{4.22}$$

sendo:

$\alpha_t(i)$ fornece a probabilidade de estar no estado q_i no instante t;

$a_{ij}b_j(O_{t+1})$ calcula o número de transições a um estado q_j emitindo o símbolo O_{t+1} a partir desse estado;

$\beta_{t+1}(j)$ conta o número de ocupações do estado q_j no instante $t+1$ e completando em seguida a seqüência de observações.

Essa probabilidade deve ser calculada sobre todos os instantes aos quais é possível de realizar uma transição, isto é, de $t=T-1$ para $t=1$.

c) Obtenção dos Parâmetros do Modelo

Para calcular o denominador da equação (4.12) recorda-se que:

$$\begin{aligned} \alpha_t(i) &= P(O_1 O_2 \cdots O_t, q_t = i / \lambda) \\ \beta_t(i) &= P(O_{t+1} O_{t+2} \cdots O_T / q_t = i, \lambda) \end{aligned} \tag{4.23}$$

Segue-se que $\alpha_t(i)\beta_t(i)$ é a probabilidade de ocupar o estado q_i ao instante t no modelo λ. Portanto, o denominador da equação (4.12) escreve-se:

$$P(\text{Transição do estado } q_i \text{ ao estado qualquer}, O/\lambda) = \sum_{t=1}^{T-1} \alpha_t(i)\beta_t(i) \tag{4.24}$$

e a equação completa pode ser escrita como:

$$a_{ij}^{NM} = \frac{\sum_{t=1}^{T-1} \alpha_t(i)\beta_{t+1}(j) b_j(v_{t+1}) a_{ij}}{\sum_{t=1}^{T-1} \alpha_t(i)\beta_t(i)} \tag{4.25}$$

Um raciocínio similar leva a escrever a equação (4.13) sob a forma:

$$b_{jk}^{NM} = \frac{\sum_{t \in O_t = v_t}^{T-1} \alpha_t(j)\beta_t(j)}{\sum_{t=1}^{T-1} \alpha_t(j)\beta_t(j)} \tag{4.26}$$

O somatório do numerador da equação (4.26) se lê: ”somar sobre todos os t para os quais O_t é o símbolo v_k”. Finalmente, a reestimação de π é dada por:

$$\pi_i^{NM} = \frac{\alpha_1(i)\beta_1(i)}{P^{BW}} \tag{4.27}$$

A modelagem das 39 palavras do léxico dos cheques bancários brasileiros segue o MEM proposto pela Figura 4.13. A topologia em questão caracteriza-se por um modelo do tipo discreto, esquerda-direita (*left-rigth* ou *Bakis Topology*) e com passo

igual a 2. Isto, deve-se ao fato de que as palavras são assim melhor modeladas e não se caracterizam, por exemplo, como modelos ergóticos, onde as transições para cada estado está na relação 1:N, onde N é o número de estados que compõem o modelo.

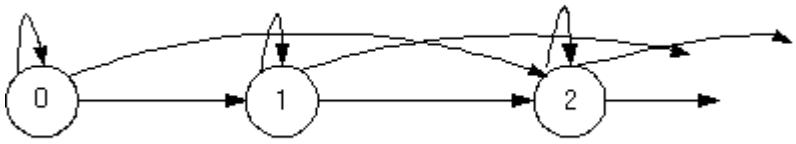

Figura 4.13: Topologia MEM para cada uma das palavras

Assim, o treinamento dos modelos MEM seguindo-se uma abordagem Global deve encontrar respostas às seguintes questões:

- Qual o número de estados adequado para cada um dos modelos-palavras?
- As seqüências de observações (extração de primitivas) são adequadas à representação das palavras?
- Os conjuntos de seqüências para treinamento atendem às necessidades inerentes desta etapa?

Com base no tamanho médio das seqüências de observações estabeleceu-se uma estratégia de busca do número de estados adequado ao modelo durante a fase de treinamento. Busca-se por um modelo que tenha de 3 ao tamanho mínimo das seqüências de observações.

Considera-se que 3 é o número mínimo de estados que um modelo pode ter seguindo um topologia esquerda-direita com passo 2. Por outro lado, sabe-se que modelos com número de estados maior do que o tamanho mínimo das seqüências de observações não conseguem ser treinados por falta de observações, não proporcionando a propagação das probabilidades no modelo.

A estratégia de busca do número adequado de estados dos modelos é auxiliada pelo processo de Validação Cruzada que permite, a cada iteração com a base de dados de treinamento, validar o modelo estimado através da base de dados de validação.

Sabe-se de [RABINER & JUANG, 1983] que seqüências de observações de tamanho muito grande fazem com que o número de estados nos MEM também sejam elevados, gerando problemas nos MEM durante a fase de treinamento. O problema mais

comum é que as probabilidades de transição entre estados e de observação de símbolos tendam a zero.

A topologia adotada para os modelos evita que as probabilidades tendam a zero, uma vez que os modelos não são do tipo ergótico. A cada estado tem-se a possibilidade de transitar para no máximo dois estados seguintes ou permanecer em si próprio. Portanto a matriz A é inicializada de maneira a atender esta especificação.

Outro detalhe da topologia é que o último estado não permite o auto-ciclo, ou seja, não se permite a permanência neste estado. Assim, o modelo não converge rapidamente para o último estado permanecendo aí até o fim das observações. Evita-se que o último estado interfira de maneira prejudicial no treinamento do modelo.

Outro fato positivo, é que os alfabetos de símbolos não são de grandes dimensões e foram validados anteriormente ao uso na etapa de treinamento.

Observa-se que as palavras com conjuntos menores de amostras realizam menos iterações, ou seja, o processo de aprendizado não é muito longo. Assim, a tendência é que a diferença entre as probabilidades de treinamento e validação seja maior. Deste modo, estes conjuntos geram maiores erros no reconhecimento, quando comparado com os demais conjuntos.

Os MEM tem a capacidade de absorver a variabilidade da escrita cursiva humana, conforme já mencionado, para tal necessita de um conjunto de dados de treinamento proporcional a esta variabilidade. Então, quanto maior a variabilidade, maior deverá ser o conjunto de dados na fase de treinamento dos modelos. Sabe-se que para as palavras do conjunto da dezena ("dez", "onze", "doze", "treze", "quatorze", "quinze", "dezesseis", "dezessete", "dezoito" e "dezenove") apresentam poucos exemplares, compondo 3,25% da base de treinamento, e portanto, sendo o processo de treinamento dos respectivos modelos prejudicado em relação às demais palavras.

O treinamento dos modelos considera o algoritmo de Baum-Welch realizando o processo de Validação Cruzada sendo a inicialização das matrizes A, B e π de acordo com as seguintes especificações:

- matriz A (matriz de transição entre estados): A[i][j], onde i = 1,..., n° de estados e j = 1,..., n° de estados, sendo A[i][j] = 1/ n° de estados,

- matriz B (matriz de probabilidade transição de observação de símbolos condicionada aos estados): B[i][k], onde i = 1,..., n° de estados e j = 1,..., n° de símbolos, sendo B[i][k] = 1/ n° de símbolos,
- vetor π (matriz de probabilidade inicial): π[i] para i = 1,..,n° de estados, onde π[1] = 1, indicando que há somente uma entrada no modelo que ocorre pelo primeiro estado.

Para [RABINER & JUANG, 1983] a inicialização das matrizes A, B e π não é um processo fácil e muito menos existe um único caminho a ser seguido. A razão para inicializar as matrizes A e B como equiprováveis é dar a chance para que as transições entre estados e as transições de observação de símbolos tenham condições de evoluir na etapa de treinamento. A inicialização randômica das matrizes pode favorecer um determinado estado ou transição de observação de símbolos prejudicando o treinamento dos modelos, tendo sido este fato observado como experimento na prática.

2.4.5.4. A Validação no Treinamento: Algoritmo de Validação Cruzada

O treinamento dos modelos considera o algoritmo de Baum-Welch [HUANG et al., 1990] e [RABINER & JUANG, 1993] associado a um processo de controle do treinamento através do algoritmo de Validação Cruzada ou *Cross-Validation*.

Os objetivos do processo de Validação Cruzada são o de monitorar o desempenho geral durante o processo de treinamento dos modelos e permitir a parada do processo de treinamento quando a melhoria da reestimação dos parâmetros do modelo não é mais possível. Este processo é realizado sobre dois conjuntos de dados: treinamento e validação. Portanto, o algoritmo de Validação Cruzada pode ser descrito em duas etapas, como apresentado a seguir.

Assim, aplica-se o algoritmo de Baum-Welch, sobre o conjunto de dados de treinamento e calcula-se a probabilidade do modelo de treinamento. Em seguida, aplica-se o algoritmo de *Forward* sobre o conjunto de dados de validação e também obtém-se a probabilidade do modelo. Desta forma, o processo de *Cross-Validation* compara as probabilidades obtidas com os conjuntos O^v de treinamento e O^u de validação realizando a reestimação dos modelos. O modelo final a ser considerado é aquele que possui demostra a convergência do processo de Validação Cruzada. Este processo é realizado até que um critério de erro ε seja atingido.

Início

Modelo_Inicial (A, B, π);
//Para a Base de Dados de Treinamento
Forward (cálculo dos valores de $\alpha_t(i)$);
Backward (cálculo dos valores de $\beta_t(i)$);
Reestimação_Modelo($\overline{\lambda} = \overline{A}, \overline{B}, \overline{\pi}$);
Se ($pr(O^u/\lambda)$ - $pr(O^v/\lambda) \geq \varepsilon$)
Então $\lambda = \overline{\lambda}$
Forward (cálculo dos valores de $\alpha_t(i)$);
Senão $\lambda^* = \arg\max \sum_{u=1}^{u^v} \log P(O^u \mid \lambda)$;

//Para a Base de Dados de Validação
Forward (cálculo dos valores de $\alpha_t(i)$);
Cálculo_$pr(O^v/\lambda)$;
Fim_Validação_Cruzada;

O objetivo do processo de Validação Cruzada é o de monitorar o desempenho geral durante o processo de treinamento dos modelos. A probabilidade do modelo de treinamento é crescente ao longo das diversas iterações do processo. Por outro lado, a probabilidade do modelo sobre os dados de validação acompanha a curva de desenvolvimento da treinamento até o ponto onde inicia-se o declínio da cruva de validação. Portanto, este processo compara as probabilidades de treinamento e de validação realizando a reestimação dos modelos. O modelo final do treinamento a ser considerado é aquele que possui a melhor probabilidade de validação, considerando-se que a curva de validação está no pico máximo. Todo este processo é realizado até que um critério de erro seja atingido. Para o estudo em questão o critério aplicado foi que a proporção entre as probabilidades, para as bases de dados de treinamento e validação, nos instantes *i-1* e *i* seja inferior a 10^{-3}.

2.4.5.5. O Reconhecimento: Algoritmo *Forward*

No problema do reconhecimento consideram-se dados [HUANG et al., 1990] e [RABINER & JUANG, 1993]:

- uma seqüência de observações desconhecidas e um conjunto de modelos λ_1, λ_2, ..., λ_{Nc} (N_c sendo o número de palavras ou de classes possíveis, e cada

modelo λ_i é definido por um conjunto de parâmetros (A, B, π) que lhe é próprio),

e

- uma descrição desconhecida O, composta de uma seqüência de T observações $O_1, O_2, ..., O_T$, onde cada O_i é um dos M símbolos do alfabeto $\{v_1, v_2, ...v_m\}$.

Pretende-se, então, dar uma interpretação para a descrição, determinando-se a classe de pertinência.

Para o reconhecimento necessita-se calcular a probabilidade $P(O/\lambda_i)$ que o modelo λ_i produz a sequência O para $i = 1, 2, ..., N_c$. A classe k reconhecida verifica o critério seguinte:

$$P(O/\lambda_k) = \underset{i, 1 \le i \le Nc}{\arg\max} \left(P(O/\lambda_i)\right) \tag{4.28}$$

Esta é a regra de máxima verossimilhança e para calcular essas probabilidades aplica-se o algoritmo de *Forward*, apresentado anteriormente.

A obtenção do modelo para cada uma das palavras constitui a determinação das matrizes A, B e π definidas como: matriz de transição entre estados, matriz de probabilidade transição de observação de símbolos condicionada aos estados e matriz de probabilidade inicial, respectivamente.

A Figura 4.14 exemplifica a metodologia aplicada para decisão de reconhecimento das palavras isoladas. O esquema explica como em [RABINER & JUANG, 1983] que para cada modelo {A, B, π} correspondente a cada palavra do léxico, toma-se a seqüência de observações $\boldsymbol{O}$ extraída de uma palavra conhecida $\boldsymbol{X}$ e faz-se o cálculo das probabilidades (algoritmo de *Forward*) por todos os possíveis modelos.

Seleciona-se, então, a probabilidade de maior valor (TOP1) e verifica-se qual o modelo $\boldsymbol{\lambda}$ correspondente. Desta forma, resulta que a seqüência $\boldsymbol{O}$ é reconhecida pelo modelo $\boldsymbol{\lambda}$ e, portanto, se o modelo $\boldsymbol{\lambda}$ corresponde a palavra $\boldsymbol{X}$ tem-se o reconhecimento correto da mesma. Os resultados obtidos apresentam as taxas de reconhecimento para os experimentos realizados, considerando-se as 5 melhores hipóteses de reconhecimento (TOP1,TOP2,...,TOP5).

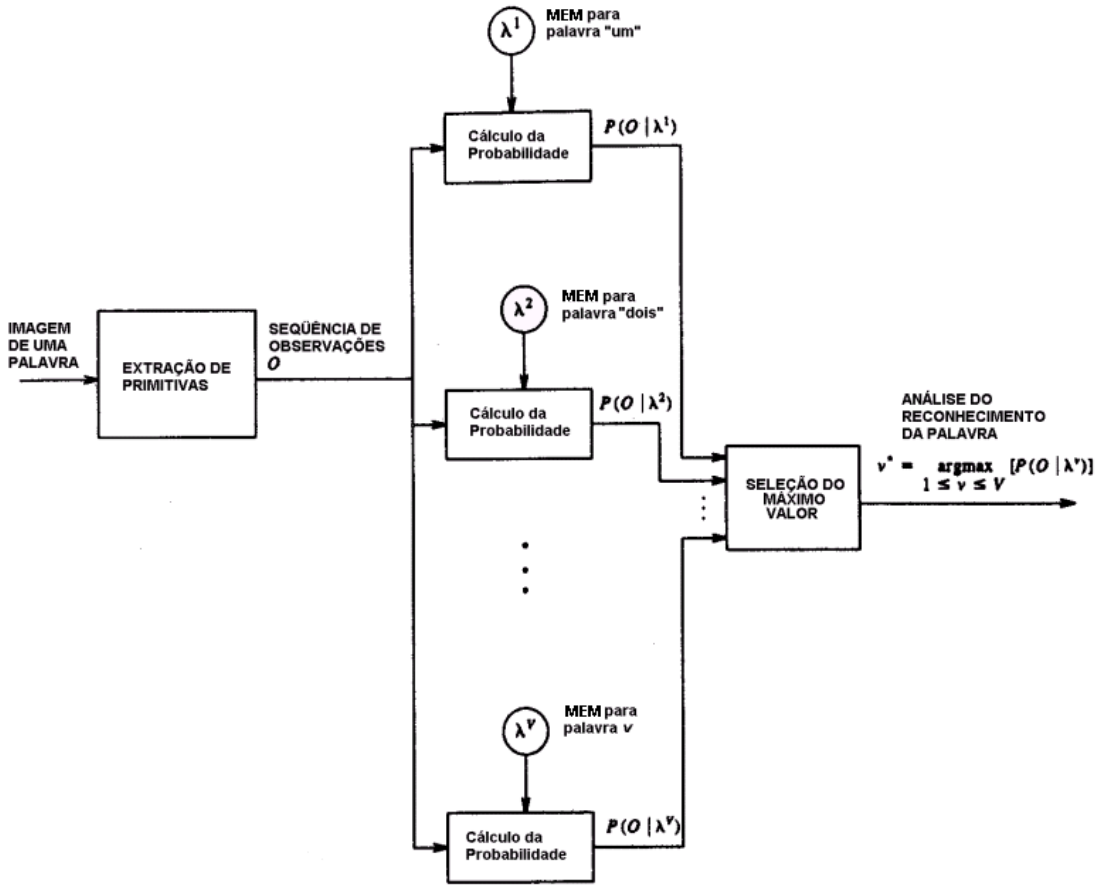

Figura 4.14: Reconhecimento de palavras isoladas MEM [RABINER & JUANG, 1983]

2.4.5.6. Vantagens e Desvantagens dos MEM

Algumas das propriedades de interesse dos MEM devem ser ressaltadas, tais como:

- Os MEM possuem uma representação complementar nos seus dois conjuntos de parâmetros A e B. As probabilidades de observação de símbolos representam a morfologia do fenômeno estudado. O outro conjunto de parâmetros, as probabilidades de transições, representa a interação temporal (ligada a ordem) entre os componentes de uma forma. Finalmente, a otimização conjunta destes desses parâmetros faz com que os MEM sejam particularmente adaptados à modelagem de sinais variantes no tempo, assim como a fala ou a escrita,

- Seja na fala ou em palavras manuscritas, o problema de reconhecimento torna-se difícil pelo fato da existência de várias fontes de variabilidade, conforme já mencionado neste capítulo. Para tratar esse problema de maneira eficaz, é necessário

uma grande quantidade de exemplos de treinamento para uma modelagem correta dos fenômenos observados. O ajuste manual entre as formas e as classes associadas, em uma base qualquer de treinamento, é portanto, uma tarefa extremamente enfadonha, na verdade impossível. Os MEM possuem esta capacidade, colocando-se inicialmente hipóteses sobre a estrutura do modelo e utilizando-se em seguida a probabilidade do modelo gerado a partir da base de treinamento (critério da máxima verossimilhança) para melhorar as estimativas de parâmetros,

- O treinamento dos parâmetros com o auxílio dos MEM e das técnicas de reestimação apóia-se sobre uma base matemática garantindo a convergência para um ponto crítico ótimo,
- O treinamento ajusta-se convenientemente ao aumento do número de exemplos da base de dados, necessitando simplesmente de um aumento linear dos cálculos,
- O reconhecimento de formas muito variáveis, tais como a fala ou a escrita, requerem a ponderação com incertezas. Os MEM sendo modelos probabilísticos, e portanto, são os candidatos naturais à representação destas formas,
- O reconhecimento implica numa pesquisa dos espaços dos estados visando uma solução ótima ou quase ótima. As pesquisas baseadas em MEM não propagam erros e permitem integrar diferentes fontes de conhecimento a cada decisão.

A utilização dos MEM pelos problemas reais encontra algumas dificuldades apresentadas a seguir:

- Os MEM se apoiam sobre a hipótese de independência das probabilidades para determinar a probabilidade de observação de uma seqüência de símbolos. Isto decorre da hipótese que as observações extraídas de um sinal são independentes entre elas, e que a geração de um símbolo de observação não é função do estado corrente da cadeia de Markov. É claro que para os sinais reais, tais como a fala ou a imagem, esta hipótese é inadequada e não pode se aproximar da realidade. Todavia, tem a vantagem de simplificar os cálculos,
- [LEVINSON et al., 1982] apud [EL YACOUBI, 1996] estudaram vários exemplos concretos para julgar o comportamento das técnicas de modelagem por MEM sobre dados ideais, e verificaram que os algoritmos de estimação de parâmetros são corretos. A lição mais importante que pode-se tirar é que mesmo sobre condições ideais (quando os dados são gerados por um MEM conhecido *a*

priori e que tem-se, portanto, a vantagem de conhecer o verdadeiro modelo), os modelos calculados sobre a base de treinamento podem não representar de maneira fiel a estrutura inerente dos dados tratados. Por consequência, uma grande atenção e validações empíricas são requeridas para uma utilização destas técnicas.

As vantagens e desvantagens da aplicação dos Modelos Escondidos de Markov possibilitam entender melhor o problema do reconhecimento e as dificuldades encontradas, principalmente, durante o processo de treinamento dos modelos.

Os experimentos realizados, bem como, os resultados obtidos e a análise de erros estão apresentados a seguir.

4.2.5.7. Experimentos Realizados

Os modelos definidos para as 39 palavras do léxico apresentam número de estados diferentes como mostrado na Figura 4.15. Observa-se que alguns modelos mantém o número de estados independentemente do conjunto de primitivas extraído.

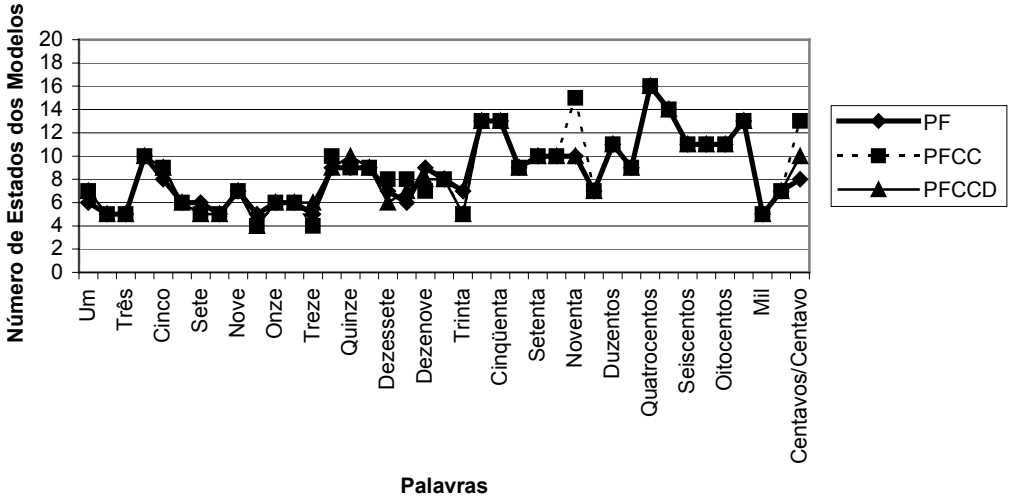

Figura 4.15: Número de estados dos modelos das palavras

As classes de palavras possuem uma similaridade em relação ao número de estados dos modelos. Isto está relacionado ao fato de que as palavras possuem uma mesma origem denunciada em seus prefixos e sufixos, além de possuírem o número de letras e, portanto, de pseudo-segmentos também similares para um mesma classe de palavras.

A Tabela 4.9 apresenta as taxas de reconhecimento para cada uma das 39 palavras do léxico estudado considerando-se os 3 conjuntos de primitivas: PP, PPCC e PPCCR.

Tabela 4.9: Taxas de Reconhecimento das Palavras para os Conjuntos de Primitivas

	Palavras	**PP (%)**	**PPCC (%)**	**PPCCR (%)**
UNIDADE	Um	80,30	83,33	**92,42**
	Dois	64,29	65,71	**72,86**
	Três	47,68	55,22	**59,70**
	Quatro	70,31	78,13	**81,25**
	Cinco	63,64	74,98	**77,27**
	Seis	35,00	50,00	**51,67**
	Sete	43,40	**51,77**	43,40
	Oito	**73,77**	**73,77**	**73,77**
	Nove	61,90	75,37	**76,19**
DEZENA	Dez	**66,67**	55,56	**66,67**
	Onze	**70,00**	40,00	50,00
	Doze	**40,00**	**40,00**	20,00
	Treze	**85,71**	57,14	71,43
	Quatorze	28,57	**42,86**	28,57
	Quinze	55,56	**77,78**	55,56
	Dezesseis	**50,00**	33,33	16,67
	Dezessete	28,57	**57,14**	28,57
	Dezoito	12,24	48,98	**66,67**
	Dezenove	55,56	70,31	**88,89**
ENTA	Vinte	52,63	**72,37**	63,16
	Trinta	**63,51**	35,14	47,30
	Quarenta	59,70	62,69	**65,67**
	Cinqüenta	78,67	**84,92**	78,67
	Sessenta	20,29	**37,67**	36,62
	Setenta	32,43	**48,65**	37,84
	Oitenta	50,00	47,22	**55,56**
	Noventa	52,38	**71,43**	65,08
ENTOS	Cem/Cento	47,35	**58,18**	**58,18**
	Duzentos	67,35	**69,39**	**69,39**
	Trezentos	31,91	46,81	**53,19**
	Quatrocentos	58,14	**69,07**	67,44
	Quinhentos	**81,25**	78,23	79,17
	Seiscentos	28,87	39,98	**44,19**
	Setecentos	55,10	**57,67**	57,14
	Oitocentos	54,00	**70,80**	66,00
	Novecentos	76,60	68,09	**80,85**
PC	Mil	59,92	83,27	**84,82**
	Reais/Real	44,14	**68,65**	67,46
	Centavos/Centavo	80,83	**86,30**	83,03
	Média	**57,22**	**67,34**	**67,66**

Observa-se que as taxas de reconhecimento são crescentes para os 3 conjuntos de primitivas em um total de 15 palavras, a saber: “um”, “dois”, “três”, “quatro”, “cinco”, “seis”, “nove”, “dezoito”, “dezenove”, “quarenta”, “oitenta”, “trezentos”, “seiscentos”, “novecentos” e “mil”.

Para a palavra “oito” a taxa de reconhecimento independe do conjunto de primitivas extraído. As palavras “doze”, “cem/cento” e “duzentos” obtiveram taxas iguais para os conjuntos PP e PPCCR.

As palavras “dez”, “onze”, “treze”, “dezesseis”, “trinta” e “quinhentos” obtiveram uma maior taxa de reconhecimento para o conjunto PP. As 14 palavras restantes obtiveram a maior taxa com o conjunto PPCC.

Desde modo, nota-se que o conjunto PPCCR (67,66%) não traz grande contribuição quando comparado ao conjunto PPCC (67,34%). Porém, ao compará-lo com o conjunto PP (57,22%) percebe-se um aumento significativo da taxa de reconhecimento. O mesmo ocorre entre os conjuntos PP e PPCC.

Pode-se concluir num primeiro instante que os conjuntos PPCC e PPCCR são similares e que o esforço computacional de rotulação dos pixels em busca de uma melhor representação das letras “C”, “S”, “E” e “Z”, e ainda, “u”, “n”, “r” e “v” não corresponde ao esperado. A aplicação do conjunto de primitivas PPCC incluindo somente a extração das concavidades e convexidades no corpo das palavras sobre a imagem do esqueleto é suficiente para melhorar a performance do método. A Tabela 4.10 apresenta os resultados obtidos com as 5 melhores hipóteses de reconhecimento das palavras isoladas (TOP1, TOP2,..., TOP5). Outros autores apresentam seus resultados considerando 10, 16 ou até 20 hipóteses para léxicos menores que 39 palavras [GUILLEVIC, 1995], [AVILA, 1996] e [KNERR et al., 1996].

Tabela 4.10: Taxa de Reconhecimento dos Conjuntos de Primitivas Perceptivas

Conjunto de Primitivas	**TOP1**	**TOP2**	**TOP3**	**TOP4**	**TOP5**
PP	57,22	73,16	80,32	84,78	88,10
PPCC	67,34	79,97	85,43	88,87	91,12
PPCCR	67,66	80,42	86,65	89,94	92,21

Outro experimento realizado com as palavras considera a probabilidade de ocorrência de cada uma das palavras na base de dados de treinamento. Assim,

$$k = \arg\max_{i} [pr(O / \lambda_i) \,.\, p(\lambda_i)] \qquad (4.29)$$

onde k é o índice i que maximiza a função (i = 1,..,N_c), N_c é o número de palavras do léxico (39 palavras), O é a seqüência de observações da palavra e $p(\lambda_i)$ é a freqüência relativa da palavra i na base de dados de treinamento. Realiza-se este experimento para considerar da freqüência de ocorrência na base de dados de treinamento de cada palavra sobre a probabilidade calculada para a palavra desconhecida e, posteriormente, computar as hipóteses de reconhecimento. Os resultados obtidos estão apresentados na Tabela 4.11.

Verifica-se um acréscimo da taxa de reconhecimento para os 3 conjuntos de primitivas, indicando que a freqüência de ocorrência das palavras na base de treinamento deve ser levada em conta no momento da formulação das hipóteses de reconhecimento.

Tabela 4.11: Taxa de Reconhecimento dos Conjuntos de Primitivas Perceptivas para Equação (4.29)

Conjunto de Primitivas	**TOP1**	**TOP2**	**TOP3**	**TOP4**	**TOP5**
PP	61,52	76,34	83,24	86,85	90,27
PPCC	69,70	79,94	85,42	88,87	91,11
PPCCR	70,61	82,44	88,08	90,53	92,84

4.2.5.8. Análise de Erros

A análise de erro é realizada sobre a matriz de confusão das palavras da base de teste. Observa-se nesta matriz a confusão existente no reconhecimento de uma palavra por outra de uma mesma classe ou por outra palavra de classe diferente. Isto, auxilia a compreensão da similaridade das formas, das dificuldades do processo de treinamento e, por conseqüência, dos processos de extração de primitivas.

Inicialmente, apresenta-se a matriz de confusão para cada um dos conjuntos de primitivas e, posteriormente, realiza-se uma análise resumo identificando os erros ocorridos intra-classe e inter-classes.

A Tabela 4.12 apresenta a matriz de confusão para o conjunto de primitivas PP. As palavras estão indicadas pelos valores numéricos para facilitar a apresentação da matriz, sendo as letras M, R e C representação das palavras-chaves: “mil”, “ reais/real” e “centavos/centavo”, respectivamente. As Tabelas 4.13 e 4.14 apresentam as matrizes de confusão para os conjuntos PPCC e PPCCR de primitivas.