Technical University of Denmark

DTU

Coding with partially hidden Markov models

Forchhammer, Søren; Rissanen, J.

Link back to DTU Orbit

DTU Library
Technical Information Center of Denmark

Coding with Partially Hidden Markov Models

Søren Forchhammer [*]

Institute of Telecommunication, 343

Technical University of Denmark

DK-2800 Lyngby, Denmark

Jorma Rissanen

IBM Research Division

Almaden Research Center, K52/802

San Jose, CA 95120-6099

Abstract: Partially hidden Markov models (PHMM) are introduced. They are a variation of the hidden Markov models (HMM) combining the power of explicit conditioning on past observations and the power of using hidden states. (P)HMM may be combined with arithmetic coding for lossless data compression. A general 2-part coding scheme for given model order but unknown parameters based on PHMM is presented. A forward-backward reestimation of parameters with a redefined backward variable is given for these models and used for estimating the unknown parameters. Proof of convergence of this reestimation is given. The PHMM structure and the conditions of the convergence proof allows for application of the PHMM to image coding. Relations between the PHMM and hidden Markov models (HMM) are treated. Results of coding bi-level images with the PHMM coding scheme is given. The results indicate that the PHMM can adapt to instationarities in the images.

1   Introduction

Hidden Markov models are powerful models with a rich mathematical structure. The hidden Markov model (HMM) is a doubly embedded stochastic process. The applications have often involved estimating the probabilities of hidden states in e.g. recognition problems. In this paper, the emphasis is not so much on recognition as on the basic problem of obtaining a good probability estimate of an observed sequence. Combined with entropy coding, this may translate directly into the number of bits to code the observation sequence. In a universal coding scheme [Ris83][Ris86], parts of the past (the context) is used to obtain predictive estimates of the current value of an observation string. This has been applied with success to e.g. bi-level images. In this paper, a combination of the HMM and the context approach is presented. The new model is called a partially hidden Markov model (PHMM). One approach to source coding is two-part coding where the coded data stream consists of an encoding of the model parameters along with the coded data. A general coding scheme along these lines is presented. In Section 2, hidden Markov models and earlier work on coding with HMM is discussed. A statement of problem is given. In Section 3, the new partially hidden Markov model is introduced. In Section 4, the new general coding scheme is presented along with a set of reestimation formulas. In Section 5, constrained versions are given. The relations to a HMM models of larger order is given. In Section 6, the convergence of the PHMM reestimation formulas is proven. The results of experiments conducted with bi-level images is given in Section 7.

---

[*]Most of the work was done while the author was on sabbatical leave at IBM Almaden Research Center.

2   Background and statement of problem

A good tutorial on hidden Markov models is given in [Rab89]. The hidden Markov model is a doubly embedded stochastic process, where one of them is not directly observable (it is hidden). The output is produced by a set of stochastic processes connected with the states of the hidden process. A formal description of the HMM and the reestimation formulas will appear as a special case of the partially hidden Markov model presented later.

For known model and parameters (known to sender and receiver), equations for sequential calculation of probabilities for coding was presented by Ott [Ott67]. Universal coding of hidden Markov sources, i.e. HMM (with unknown order and parameters) was presented in [LN94]. Shtarkov [Sht92] presented a universal coding scheme for switching sources (with unknown parameters) which include the HMM as a special case. Unfortunately, both these universal coding schemes involve summing over exponentially many sequences involving hidden states. Even though the sequences may be grouped resulting in only polynomially complexity in the data length, from a practical point of view this is still prohibitive.

In the next section, our modification to the HMM is introduced. It is similar to a switching source in the sense that the outputs from a state(/component) are not independent. The difference is that for the PHMM, the current state and outputs may be dependent on previous outputs from different states(/components).

The aim is to develop a coding scheme of reasonable complexity (linear in the data length) which uses hidden states with unknown parameters.

3   Partially Hidden Markov Models

In this section, we combine the hidden Markov model with an ordinary Markov model. The new model formulation is called a partially hidden Markov model. There are N (hidden) states, each denoted $\bar{s}_i$. The produced observation sequence of length T is denoted $O = O_1, O_2...O_T$. The output symbols $O_t = k$ belong to a discrete alphabet $\{1,2...N_k\}$.

The idea is to combine hidden states $\bar{s}$ with observable states s. A PHMM is a process $\{O^t, \bar{q}^t\}$ defined by the axioms and parameters (the right-hand sides)

$$P(O_t|O^{t-1}, \bar{q}^t) = P(O_t|q_t, \bar{q}_t) \tag{1}$$

$$P(\bar{q}_t|O^{t-1}, \bar{q}^{t-1}) = P(\bar{q}_t|q_{t-1}, q_t, \bar{q}_{t-1}) \tag{2}$$

where q_t is a mapping of O^{t-1} with a finite range labelled $\{1,2...K\}$. q_t could e.g. be a subset of O^{t-1} defined by a fixed template. This means that q_t combined with the hidden state $\bar{q}_t$ conditions the output O_t. (q_t, q_{t-1}) conditions the state transition between hidden states $\bar{q}_{t-1}$ and $\bar{q}_t$. The (partially) hidden Markov model is described by the parameter set $\lambda = (\pi, A, B)$, where π gives the probabilities of the initial hidden state, A the conditional hidden state transition probabilities a_{ij}, and B the conditional output probabilities b_j for each of the states.

In order to reduce the number of parameters, a context formulation of the PHMM may be used. When calculating the probabilities for (O_{t+1}), part of the causal data is used as context for conditioning the transition and the output probabilities. Let I denote mappings of the observations. One context $(v_t = I_{v_t}(O^t))$ is used for conditioning

the transition from the hidden state $\bar{q}_t$ to $\bar{q}_{t+1}$. Another context $(w_{t+1} = I_{w_{t+1}}(O^t))$ is combined with the hidden state $\bar{q}_{t+1}$ for conditioning the probability of output $(O_{t+1} = k)$. v_t and w_{t+1} assume one of N_l and N_n values, respectively. Hence, the parameters of the partially hidden Markov model $\lambda = (\pi, A, B)$ consist of the following

$$\pi_i = P(\bar{q}_1 = \bar{s}_i) \tag{3}$$

$$a(i, j, l) = P(\bar{q}_{t+1} = \bar{s}_j | v_t = s_l, \bar{q}_t = \bar{s}_i) \tag{4}$$

$$b_j(k, n) = P(O_{t+1} = k | w_{t+1} = s_n, \bar{q}_{t+1} = \bar{s}_j). \tag{5}$$

where s_l and s_n are context values. A shorter notation is obtained when t is used to index the parameters:

$$a_t(i, j) = P(\bar{q}_{t+1} = \bar{s}_j | v_t, \bar{q}_t = \bar{s}_i) \tag{6}$$

$$b_{t+1}(j, k) = P(O_{t+1} = k | w_{t+1}, \bar{q}_{t+1} = \bar{s}_j). \tag{7}$$

Given the model parameters and the observation sequence until t, O^t, we may sequentially calculate the probability of being in a specific hidden state $\bar{s}_i$ at t,

$$\alpha_t(i) = P(O^t, \bar{q}_t = \bar{s}_i | \lambda). \tag{8}$$

The definition of $\alpha_t(i)$ as written above is the same as for the HMM [Rab89], although the influence of O^t is changed, as reflected in the induction formulas

$$\alpha_1(i) = \pi_i b_i(O_1), \qquad 1 \leq i \leq N \tag{9}$$

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_t(i, j) \right] b_{t+1}(j, k), \quad 1 \leq t \leq T - 1, \quad 1 \leq j \leq N \tag{10}$$

If the seen parts of the states can assume only one fixed value, we have the formulation of the HMM given in e.g. [Rab89].

4   A general coding scheme employing PHMM

A general two part coding scheme based on the PHMM is presented in this section. In the code stream, the model parameters are sent along with the coded data.

4.1   Coding with Partially Hidden Markov Models

The partially hidden Markov model may be used to encode data. The probability of the observed sequence until t; i.e., O^t, given the PHMM model parameters is given by summing the forward variables at t, $P(O^t | \lambda) = \sum_{i=1}^{N} \alpha_t(i)$. Combining with (10) we obtain a term for sequential calculation of the ideal codelength

$$P(O_{t+1} | O^t, \lambda) = \frac{P(O^{t+1} | \lambda)}{P(O^t | \lambda)} = \frac{\sum_{j=1}^{N} \left[\sum_{i=1}^{N} \alpha_t(i) a_t(i, j) \right] b_{t+1}(j, O_{t+1})}{\sum_{i=1}^{N} \alpha_t(i)}. \tag{11}$$

The term $\frac{\sum_{i=1}^{N} \alpha_t(i) a_t(i,j)}{\sum_{i=1}^{N} \alpha_t(i)}$ is a mixture weighting expressing the probability of state $\bar{q}_{t+1}$ $= \bar{s}_j$, given the observation sequence O^t up to t. This is also a renormalization of

the α values. The ideal code length is given by $ln[P(O|\lambda)] = \sum ln[P(O_{t+1}|O^t, \lambda)]$. An actual coding may be performed by applying arithmetic coding to the sequence of conditional probabilities. For bi-level output data, $P(O_{t+1}|O^t, \lambda)$ is sufficient. To use this for encoding, the model has to be given beforehand, or otherwise a codeable model is sent along with the coded data. A simple solution to quantization and encoding PHMM parameters is given later.

4.2 Reestimation of parameters

If the model parameters are not given, we have the problem of estimating the parameters, including those involving the hidden states. For a given model order of a HMM, the Baum-Welch method iteratively converges towards a local maximum of $P(O|\lambda)$ over the parameters λ [LRS83]. The model parameters are reestimated after each pass of the data set. The reestimation formulas are modified for the PHMM as specified below.

Given the entire observation sequence and the model parameters λ, the forward variable may be supplemented with a backward variable to obtain the probability of being in a given hidden state $\bar{s}_i$ at time t, $\gamma_t(i) = P(\bar{q}_t = \bar{s}_i|O, \lambda)$.

For the HMM, the backward variable at t for state $\bar{s}_i$, $\beta_t(i)$ is defined as $P(O_{t+1}^T|\bar{q}_t = \bar{s}_i, \lambda)$ [Rab89]. For the HMM, we further have [Rab89]

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^N \alpha_t(i)\beta_t(i)}. \tag{12}$$

For the PHMM, $\gamma_t(i)$ is still given as above. The backward variable β is now defined by solving Eq. 12.

$$\beta_t(i) = \frac{P(\bar{q}_t = \bar{s}_i|O, \lambda)P(O|\lambda)}{\alpha_t(i)}. \tag{13}$$

This way β involves the conditional probabilities of (6 and 7) and is rather a completion of the forward α-pass than a probability in its own right.

The induction formula determining the backward variable β becomes

$$\beta_T(i) = 1, \qquad 1 \le i \le N. \tag{14}$$

$$\beta_t(i) = \sum_{j=1}^N a_t(i, j)b_{t+1}(j, k)\beta_{t+1}(j), \quad t = T-1, T-2, ..., 1, \ \ 1 \le i \le N. \tag{15}$$

Introducing the contexts v_t and w_{t+1}, and thereby part of O^t, as given by a_t and b_{t+1} in the definition of β may be viewed as a look 'ahead' in the backward β pass. There is no conflict in this as β will only be used in conjunction with α values which hold the information outside of O_{t+1}^T of the contexts used by β. Using the HMM definition of β would require the part of the contexts involved in $\beta_t(i)$ belonging to O^t to be incorporated in the hidden states. (The reason for this is given in Section 6.2.) Therefore, introducing the conditioning on part of O^t in β_t gives a much smaller hidden state space and immediately leads to fewer parameters and efficient implementation.

The reestimation expression of π is the same as for the HMM although the calculations of the involved variables have changed;

$$\bar{\pi}_i = \gamma_1(i) \tag{16}$$

For A the reestimation formula is

$$\bar{a}(i,j,l) = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)c_{a,t}(l)}{\sum_{t=1}^{T-1} \gamma_t(i)c_{a,t}(l)}, \tag{17}$$

where

$$\xi_t(i,j) = \frac{\alpha_t(i)a_t(i,j)b_{t+1}(j,k)\beta_{t+1}(j)}{P(O|\lambda)}, \tag{18}$$

and $c_{a,t}(l)$ is 1 if $v_t = s_l$ and 0 otherwise. For B, the reestimation formula is

$$\bar{b}_j(k,n) = \frac{\sum_{t=1}^{T} \gamma_t(j)c_{b,t}(n)d_t(k)}{\sum_{t=1}^{T} \gamma_t(j)c_{b,i}(n)}, \tag{19}$$

where $c_{b,t}(n)$ is 1 if $w_{t+1} = s_n$ and 0 otherwise and $d_t(k)$ is 1 if $O_t = k$ and 0 otherwise.

If the seen parts of the states can assume only one fixed value, we have the reestimation formulation for the HMM given in e.g. [Rab89].

The PHMM definition of the contexts, v_t and w_{t+1}, are just given by general mappings $I(O^t)$ of the causal part of the observations. One choice is to use a template and let the values within the template define the contexts. Where the template extends beyond the observation boundary, conditions have to be specified e.g. by zero-padding. Relating the observable part of the states to the contexts of the two PHMM formulations, the context of the transitions v_t is the union of q_t and q_{t+1}. The context w_{t+1} of the outputs is q_{t+1}. An example of a PHMM state $(q_t, \bar{q}_t)$ is that the hidden part is part of the future data of the observation sequence and the seen part is part of the past, both given by templates (Fig. 1).

4.3   The general coding scheme

As shown previously, we may code the data in a forward pass given the model parameters λ. The reestimation formulas given by (16, 17 and 19) may be used to improve the model parameters if they are not already at a local optimum. The only question left is initialization. If the hidden states are (mappings of) part of the future data, the model may be initialized by counting over the data set to be coded. Any finite representation of the parameters may be used. A slight modification of a uniform scalar quantizer was used. Initially, the parameters of the PHMM model $\lambda = (A, B, \pi)$ were quantized to a fixed precision for each of parameter groups A, B and π. The probabilities were therafter adjusted to add up to one (with no negative probabilities) making in unneccessary to code the last parameter. Furthermore, precautions were taken to ensure that the probability to be coded is non-zero.

The general coding scheme for a given model order is thereafter:
1) Initialization of the model parameters $\lambda = (\pi, A, B)$ by counting.
2) Reestimation n times of π (by 16), A (by 17) and B (by 19).
3) Encoding of the reestimated PHMM parameters, e.g. coding quantized parameters.
4) (Arithmetic) coding obased on $P(O_{t+1}|O^t, \lambda)$ (11) in a forward pass.

As for the HMM, the PHMM calculations including the reestimations may be effectively implemented based on a trellis structure with the hidden parts of the states as nodes. For a fixed number of iterations the complexity is linear in the data length. Furthermore no training set is required due to the simple initialization rule above.

5   Constraints and parameter tying

To reduce the number of parameters and to introduce certain structures, the transitions may be constrained, e.g. by tying parameters. The use of Moore [Rab89] or Mealy [JM80] machines for the HMM are examples of model choices with different structures.

To reduce the number of (free) parameters to be estimated, the approach of tying parameters has been developed [JM80]. The tying was defined in conjunction with a Mealy machine formulation of the HMM, where the output is given by the transition. The state $\bar{s}_1$ is tied to $\bar{s}_2$ if there is a 1-1 correspondence between their outgoing (Mealy) transitions with corresponding probabilities being equal [JM80].

The transitions between the hidden states may be constrained. One instance of this is that two consecutive hidden states are partly overlapping, i.e. only a subset of $\bar{q}_{t+1}$ values is possible, given $\bar{q}_t$. Using a template to specify hidden state values (for initialization), the overlap between two consecutive hidden templates gives the overlap constraint. For images, this may be relaxed just imposing the constraint on a line by line basis.

6   Convergence and equivalence

The reestimation of the partially hidden Markov model given previously still converges to a local maximum of $P(O|\lambda)$ over λ. The proof of the convergence of the HMM [LRS83] applies with a few modifications. The proof evaluates the probabilities of the observations $P(O|\lambda)$ and $P(O|\bar{\lambda})$ conditioned on two consecutive sets of parameters in the reestimation process. This is done by summing over all sequences (of hidden states). This still applies if we just consider the hidden part of the PHMM states. Evaluating the above probabilities the terms are organized according to the state transitions and the observed symbols. This grouping is extended to include the seen conditioning contexts for the PHMM.

6.1   Proof of the convergence of reestimation formulas

Let S denote the number of hidden state sequences $\bar{q}^T$ and s be a specific state sequence. Let u_s and v_s be defined as $P(\bar{q}^T = s, O|\lambda)$ and $P(\bar{q}^T = s, O|\bar{\lambda})$, respectively. The conditional probabilities of the observation, given the current model λ and the new estimated parameters $\bar{\lambda}$, respectively, are compared based on the following inequality [LRS83],

$$D(\lambda, \bar{\lambda}) = \ln \frac{P(O|\bar{\lambda})}{P(O|\lambda)} \geq [Q(\lambda, \bar{\lambda}) - Q(\lambda, \lambda)] \tag{20}$$

where $Q(\lambda, \bar{\lambda}) \equiv \sum_s u_s \ln v_s$.

Inserting the expressions for u_s and v_s and regrouping terms according to the parameters conditioned on their contexts, it can be seen that

$$Q(\lambda, \bar{\lambda}) = \sum_{l=1}^{N_l} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} c_{ijl} \ln \bar{a}_{ijl} + \sum_{n=1}^{N_n} \sum_{j=1}^{N_i} \sum_{k=1}^{N_k} d_{jkn} \ln \bar{b}_{jkn} + \sum_{i=1}^{N_i} e_i \ln \bar{\pi}_i, \tag{21}$$
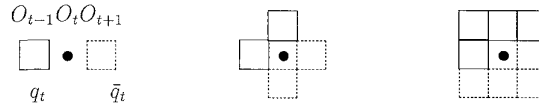
$$O_{t-1}O_tO_{t+1}$$

Figure 1: Left) The seen part of the state is the previous value, $q_t = O_{t-1}$ and the hidden part is the hidden value $\bar{q}_t = O_{t+1}$. Right) Templates defining states $(q_t, \bar{q}_t)$ of a PHMM.

where N_l is the number of A-contexts for a_{ij}, N_n is the number of B-contexts and N_i and N_j are the number of predecessor i and successor j hidden states, while N_k is the size of the output alphabet. Further, summing over all S hidden state sequences

$$c_{ijl} = \sum_s P(s, O|\lambda) n_{ijl}(s) \tag{22}$$

$$d_{jkn} = \sum_s P(s, O|\lambda) m_{jkn}(s) \tag{23}$$

$$e_i = \sum_s P(s, O|\lambda) f_i, \tag{24}$$

where for sequence s, $n_{ijl}(s)$ is the number of transitions from $\bar{s}_i$ to $\bar{s}_j$ given $v = s_l$, $m_{jkn}(s)$ is the number of times symbol k is generated in state $\bar{s}_j$ given $w = s_n$, and f_i is 1 if initial state is $\bar{s}_i$. Equation 21 for $Q(\lambda, \bar{\lambda})$ consists of $(N_l + N_n)N_i + 1$ independent sets of terms with the constraint $\sum_i x_i = 1$. Maximizing $\sum_i c_i \ln x_i$ gives exactly the reestimation formulas.

Generally speaking, it is only required that there is a well-defined sequence of contexts (v_t, w_{t+1}), which are mappings of the causal part (O^t) of the observations for the given traversal. This context sequence must be invariant to the reestimations of parameters but it may vary with t. Later this will be used in conjunction with the border problems of an image, using slightly different mappings at the borders.

6.2 Equivalence conditions for HMM and PHMM

Consider a subset of PHMM models, where the context mappings are invariant with respect to t and the dependence on O^t only involves the values O^t_{t-M}. For this subset, an equivalent HMM model of a larger order may be derived. Using a Mealy type HMM, two constraints are imposed on the HMM. One constraint ensures that the new state holds the information of the last output. The other constraint ensures that the new state holds information of the last M values, i.e. O^t_{t-M} in the hidden state.

7 Coding of (bi-level) images

The general PHMM coding scheme of Section 4 has been applied to bi-level images. The initialization of parameters was performed on the individual images to be coded as described previously. Therafter, the parameters were improved by applying the reestimation formula. These parameters are explicitly coded before coding the actual data.

Figure 2 depicts some of the templates used to define states. In this report, the image is traversed line by line giving a one-dimensional structure of the hidden states.

Figure 2: Left) A- and hidden template conditioning transitions to next set of hidden states. Right) B- and hidden template conditioning output probabilities.

Table 1: PHMM ideal code lengths after initialization and 8 iterations on three images

Image	0 iterations	8 iterations	Reduction
CCITT 4	514,414.8	450,267.5	12%
JBIG S06	3,534,684.7	1,932,944.9	45%
JBIG S09	733,906.3	633,482.1	14%

The contexts can introduce a two-dimensional structure. At the borders zero-padding is used. The context of the first pixel of a row is combined with the hidden states of the former row. The reestimation still converges as we have a fixed sequence of contexts which does not change with the reestimation. Thus advantage is taken of using a time-varying mapping for the contexts.

7.1 Experimental Results

The general coding scheme was applied to bi-level images. The results were compared with those of JBIG [JBI93], the best standard for coding bi-level images. Three different bi-level images, from the CCITT and the JBIG Stockholm test sets were chosen. A text page (CCITT4), an error-diffused image (JBIG, S09) and a mixture of (positive and negative) text and halftone (JBIG, S06) were used. Different model orders were tried out to find good models.

The context mappings and the initialization of the hidden states were given by templates. Initial experiments showed that A- and hidden-templates corresponding to a PHMM state $(q_t, \bar{q}_t)$ given by the 8-neighbors (Fig. 1 and Fig. 2) gave good results. The A-template specifies v_t. The hidden template specifies the number of hidden states and it is used for calculating the values used for initializing the PHMM parameters. This corresponds to a 6-pixel A-template and a 4-pixel hidden-template (Fig. 2 left). The transitions of the hidden states were furthermore overlap constrained on a line by line basis, leaving only two free pixels of the successor hidden state, $\bar{q}_j$. This basic setting of A- and hidden-templates was used in combination with various B-templates throughout the further experiments. The B-template specifies w_{t+1}.

An experiment over the 3 chosen test images is summarized in Table 1. The B-templates had 6 pixels (e.g. as in Fig. 2 right). The results indicate that the reestimation improves the compression result. The tests on S06 with mixed text and image gave the highest improvement. A likely explanation is that, the simple layout structure in S06 was efficiently captured to yield a 45% reduction of code length after just 8 iterations. (These results are without quantization.)

Further experiments were carried out for the two JBIG test images (S06 and S09). The figures for PHMM coding are obtained by adding the cost of coding the quantized

Table 2: Code lengths for JBIG images S06 and S09.

	JBIG image S06		JBIG image S09	
Code	Code length	Context size	Code length	Context size
No coding	13,369,344		1,048,576	
MMR	7,970,024		2,293,664	
JBIG	2,049,160	10	594,312	10
JBIG	1,778,256	10 (Adaptive)		
PHMM	1,599,380.3	6(+4)	589,974.5	8(+4)
PHMM	1,463,564.0	8(+4)		
PHMM	1,459,579.2	9(+4)		
2-pass	1,522,748.2	13	699,140.5	8
2-pass			607,355.4	13
Predictive	1,267,644.5	18	580,187.2	13
Predictive			553,196.6	16
Context	1,195,163.7	22	546,070.1	17

parameters to the ideal code length $ln[P(O|\lambda)]$ based on these quantized parameters. (The overhead of applying a good arithmetic coder being negligible.) For the JBIG images S06 (4352 by 3072 pels) and S09 (1024 by 1024 pels) the best PHMM results and results of other compression methods are summarized in Table 2. The current FAX MMR G4 standard did not perform well on either image. The JBIG standard gave good results. For S06, the results of using PHMM coding with a B-template of 6 and 8 pixels gave a reduction in code length of 23% and 29%, respectively, in comparison with the default JBIG. The reduction was 12% and 18%, respectively, in comparison with JBIG using the one adaptive template pixel. A B-template of 9 pixels gave a marginally better result (0.3%) than the best result with 8 B-template pixels. For the 8-pixel B-template quantization with 5, 9 and 12 bits precision for the A, B and π parameters, respectively, was used. The cost of coding the parameters was 52,404 bits. This accounts for almost 10% of the code length.

For further comparison, a two-part coding using the same quantizer as for the PHMM was applied. For this two-part code a greedy search was carried out to determine the best template pixels for conditioning the probabilities. The overall best result was obtained with 13 template pixels. The PHMM coding with 8 B-template pixels gave a slightly better (3%) result than the best two-part code described above. Using one-pass predictive coding with implicit coding of the parameters, larger templates could be used and better results obtained. The template was again determined by a greedy search. Finally, a Context coding (similar to the one in [Ris86]) was performed using 22 pixels in the tree. The pixels were chosen from the results of the greedy search above. This produced the best result.

In the tables, context sizes refer to (max.) number of pixels used for conditioning the output probability. For the PHMM, this is the number of pixels in the B-template plus the hidden template.

For S09, the best PHMM result was obtained with an 8-pixel B-template. The PHMM coding used quantization with 3, 6 and 8 bits precision for the A, B and π parameters, respectively.

The predictive codings are primarily more effective in coding the model (implicitly). The comparisons indicate that a more effective coding of the PHMM parameters, local adaption of the probabilities and adaption of the template pixels would improve the results of the PHMM. In the comparisons using the same coding of the model

parameters, the PHMM gave the best results.

8 Conclusions

Partially hidden Markov modelling, a new approach to hidden Markov modelling, has been presented. It combines the powers of hidden Markov models with that of Markov models, making explicit use of both the seen past and the hidden states. A general coding scheme based on PHMM modelling and reestimation for unknown parameters was presented. For unknown parameters, reestimation formulas for the PHMM were given and their convergence proven. The proof also holds when part of the PHMM is variant in time. This was utilized in conjunction with images and crossings at the image borders. Using the PHMM, 2-dimensional dependence is efficiently introduced in HMM for coding purposes. The reestimation formulas and the model formulation allow efficient implementation. For a fixed number of reestimations, the complexity is linear in the data length. Experiments conducted with bi-level images indicate that the PHMM can adapt to instationarities. Good results were obtained coding bi-level images. For the two chosen images of the JBIG test set, PHMM coding outperformed JBIG. For one of the images by 18% compared with the best JBIG setting. Comparison with coding schemes with implicit coding of parameters indicated areas of improvement for the PHMM coding.

References

[JBI93] JBIG. Progressive Bi-level Image Compression. ISO/IEC International Standard 11544, 1993.

[JM80] F. Jelinek and R. L. Mercer. Interpolated estimation of Markov source parameters from sparse data. In E. S. Geselma and L. N. Kanal, editors, *Pattern Recognition in Practice*, pages 381–397. North-Holland, 1980.

[LN94] C-C. Liu and P. Narayan. Order estimation and sequential universal data compression of a hidden Markov source by the method of mixtures. *IEEE Trans. Inf. Theory*, 40(4):1167–1180, July 1994.

[LRS83] S. E. Levinson, L. R. Rabiner, and M. M. Sondhi. An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition. *BSTJ*, 62(4):1035–1074, April 1983.

[Ott67] G. Ott. Compact encoding of stationary Markov sources. *IEEE Trans. Inf. Theory*, 13(1):82–86, January 1967.

[Rab89] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, February 1989.

[Ris83] J. Rissanen. A universal data compression system. *IEEE Trans. Inf. Theory*, 29(5):656–664, September 1983.

[Ris86] J. Rissanen. Complexity of strings in the class of Markov sources. *IEEE Trans. Inf. Theory*, 32(4):526–532, July 1986.

[Sht92] Y. Shtarkov. Switching discrete sources and its universal encoding. *Problemy Peredachi Informatsii*, 28(3):95–111, July-September 1992. (English translation pp. 282-296).