COMPARISON OF THE EFFECTIVENESS OF 1D AND 2D HMM IN THE PATTERN RECOGNITION

JANUSZ BOBULSKI

Institute of Computer and Information Science, Faculty of Mechanical Engineering and Computer Science, Czestochowa University of Technology, 73 Dabrowskiego Str., 42-200 Czestochowa, Poland,

januszb@icis.pcz.pl

Abstract. Hidden Markov Model (HMM) is a well established technique for image recognition and has also been successfully applied in other domains such as speech recognition, signature verification and gesture recognition. HMM is widely used mechanism for pattern recognition based on 1D data. For images one dimension is not satisfactory, because the conversion of one-dimensional data into a two-dimensional lose some information. This paper presents a solution to the problem of 2D data by developing the 2D HMM structure and the necessary algorithms.

1 Introduction

Hidden Markov models are omnipresent mechanism for data classification. They are establish technique in pattern recognition, speech recognition, character recognition, biological sequence analysis, texture analysis, face recognition, financial data processing, etc [1, 2, 3, 4, 5]. The wide HMM application is a result of their effectiveness and advantages resulting from unsupervised learning. In theory, supervised and unsupervised learning differ only in the causal structure of the model. In supervised learning, the model defines the effect one set of observations, called inputs, has on another set of observations, called outputs. In unsupervised learning, all the observations are assumed to be caused by latent variables, that is, the observations are assumed to be at the end of the causal chain. In practice, models for supervised learning often leave the probability for inputs undefined. This model is not needed as long as the inputs are available, but if some of the input values are missing, it is not possible to infer anything about the outputs. If the inputs are also modelled, then missing inputs cause no problem since they can be considered latent variables as in unsupervised learning. In unsupervised learning, the learning can proceed hierarchically from the observations into ever more abstract levels of representation. Each additional hierarchy needs to learn only one step and therefore the learning time increases linearly in the number of levels in the model hierarchy [6]. In spite of recalled virtues 1D HMM is unpractical in image processing, because the images are two-dimensional. When we convert an image from 2D to 1D, we lose some information. So, if we process two-dimensional data, we should apply two-dimensional HMM, and this 2D HMM should works with 2D data. One of solutions is pseudo 2D HMM [7, 8]. This model is extension of classic 1D HMM. There are super-states, which mask one-dimensional hidden Markov models. Linear model is the topology of superstates, where only self transition and transition to the following superstate are possible. Inside the superstates there are linear 1D HMM. The state sequences in the rows are independent of the state sequences of neighbouring rows. Additional, input data are divided to the vector. So, we have 1D model with 1D data in practise. Interesting results showed in paper [9]. This article presents analytic solution and proof of correctness two-dimensional HMM. But this 2D HMM is similar to MRF [10, 11], which works with one-dimensional data and can be apply only for left-right type of HMM. An extension of the HMM to work on two-dimensional data is 2D HMM. A 2D HMM can be regarded as a combination of one state matrix and one observation matrix, where transition between states take place according to a 2D Markovian probability and each observation is generated independently by the corresponding state at the same matrix position. It was noted that the complexity of estimating the parameters of a 2D HMMs or using them to perform maximum a posteriori classification is exponential in the size of data. Similar to 1D HMM, the most important thing for 2D HMMs is also to solve the three basic problems, namely, probability evolution, optimal state matrix and parameters estimation. This article presents real solution for 2D problem in HMM. There is showing true 2D HMM which processes 2D data. Moreover the presented algorithms are regarding ergodic models, rather than of type "left-right" [12]. This paper presents an automatic pattern recognition system which uses two dimensional wavelet transform of second level decomposition for features extraction, and the classification module bases on two dimensional hidden Markov models, which work with two dimensional data.

2 1D HMM

Hidden Markov Model allows us to specify the probability of unobserved directly sequence of events. A HMM consists of two stochastic processes. The first stochastic process is a Markov chain that is characterized by states and transition probabilities. The states of the chain are externally not visible, therefore "hidden". The second stochastic process produces emissions observable at each moment, depending on a state-dependent probability distribution (Fig. 1).

We can describe Hidden Markov Model with following parameters:

- N is the number of states.
- $Q = q_1, \dots, q_N$ is the set of states. Note that the Hidden Markov Model keeps no history, so the only thing which it can remember is what state it is in now. The states of a Hidden Markov Model are hidden; we never observe them directly.
- The number of symbols M
- $O = O_1, \dots, O_M$ is the set of symbols that may be emitted.
- $\pi \in [0, 1] = \pi_1, \dots, \pi_N$ is the initial probability distribution on the states. It gives the probability of starting in each state. We expect that $\sum_{i=1}^{N} \pi_i = 1$. We should think of π as a column vector.
- $A = (a_{ij})1 \leq i, j \leq N$ is the transition probability matrix. If the "machine" is in state j, it may be in state i on the next clock tick with probability a_{ij} . We expect that $a_{ij} \in [0, 1]$ for each i and j, and that $\sum_i a_{ij} = 1, 1 \leq i, j \leq N$ for each j.
- $B = (b_{ij})1 \leq i \leq M, 1 \leq j \leq N$ is the emission probability matrix, if the "machine" is in state j, it may emit symbol i on with probability b_{ij} .

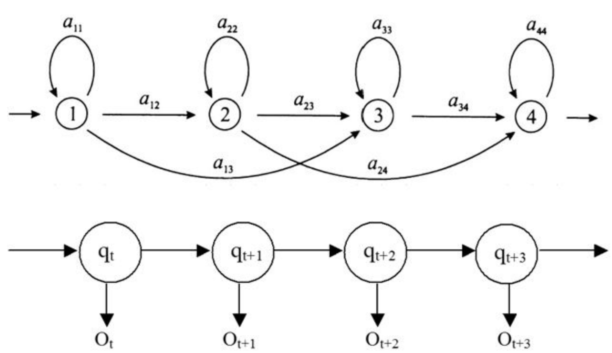

Fig. 1: One-dimensional HMM

There are two fundamental problems which must be solved in order to build the pattern recognition system:

1. Given observation $O = (o_1, o_2, \ldots, o_T)$ and model $\lambda = (A, B, \pi)$, efficiently compute $P(O|\lambda)$

2. Given observation $O = (o_1, o_2, \ldots, o_T)$, estimate model parameters $\lambda = (A, B, \pi)$ that maximize $P(O|\lambda)$

To solve the problem 1 we can use well know forward-backward algorithm, and to solve the problem 2 - Baum-Welch algorithm [1, 12]. HMM requires three probability measures to be defined, A, B, π and the notation $\lambda = (A, B, \pi)$ is often used to indicate the set of parameters of the model. The parameters of the model are generated at random at the beginning. Then they are estimated with Baum-Welch algorithm, which is based on the forward-backward algorithm. The forward algorithm calculates the coefficient $\alpha_{t(i)}$ (probability of observing the partial sequence $(o_1, \ldots, o_t)$ such that state q_t is i). The backward algorithm calculates the coefficient $\beta_{t(i)}$ (probability of observing the partial sequence $(o_{t+1}, \ldots, o_T)$ such that state q_t is i). The parameters of new model λ, based on λ_0 and observation O, are estimated from equation of Baum-Welch algorithm [1], and then are recorded to the database.

3 2D HMM

Presented solutions are for Markov model type "left-right", and not ergodic. So, we present solution to problems 1 and 3 for two dimensional data, which is sufficient to build a image recognition system. The statistical parameters of the 2D model (Fig. 2 and 3):

- The number of states of the model N^2
- $Q = q_1, \ldots, q_{N^2}$ is the set of states.
- The number of data streams k_1 x $k_2 = K$
- The number of symbols M
- Oservation sequance $O = \{o_t\}, 1 \leq t \leq M, o_t$ is square matrix simply observation with size k_1 x $k_2 = K$
- The transition probabilities of the underlying Markov chain, $A = \{a_{ijl}\}, 1 \leq i, j \leq N, 1 \leq l \leq N^2$, where a_{ij} is the probability of transition from state ij to state l
- The observation probabilities, $B = \{b_{ijm}\}, 1 \leq i, j \leq N, 1 \leq m \leq M$ which represents the probability of gnerate the m_{th} symbol in the ij_{th} state.
- The initial probability, $\Pi = \{\pi_{ijk}\}, 1 \leq i, j \leq N, 1 \leq k \leq K$.

3.1 Solution problem 1 for 2D

Forward Algorithm

- Define forward variable $\alpha_t(i, j, k)$ as:

$$\alpha_t(i, j, k) = P(o_1, o_2, \ldots, o_t, q_t = ij|\lambda) \quad (1)$$

- $\alpha_t(i, j, k)$ is the probability of observing the partial sequence $(o_1, o_2, \ldots, o_t)$ such that the the state q_t is i, j for each k_{th} strem of data

- Induction

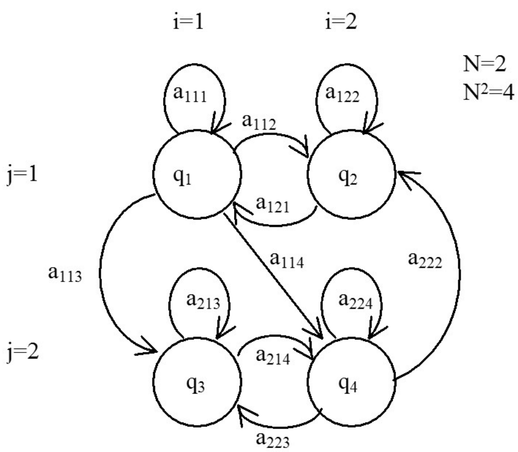

Fig. 2: Two-dimensional ergodic HMM

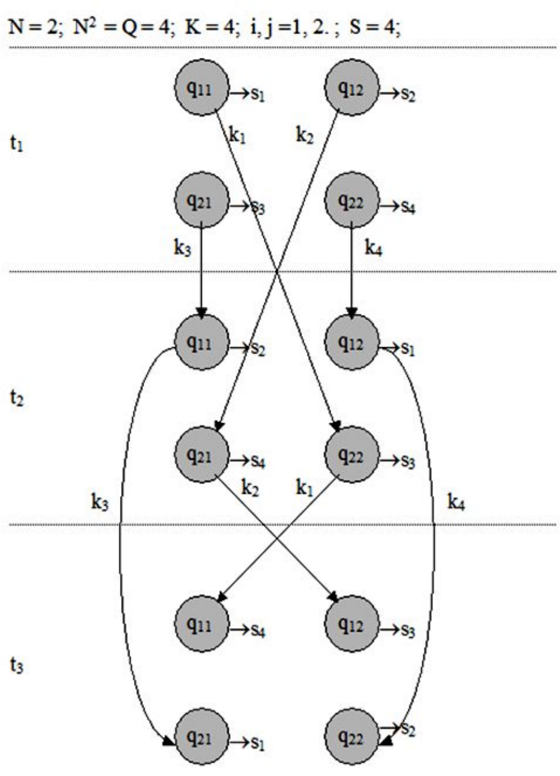

Fig. 3: Two-dimensional HMM in in subsequent time steps ($s_1 - s_4$ - observations)

1. Initialization:

$$\alpha_1(i,j,k) = \pi_{ijk} b_{ij}(o_1) \tag{2}$$

2. Induction:

$$\alpha_{t+1}(i,j,k) = \left[\sum_{l=1}^{N} \alpha_t(i,j,k) a_{ijl}\right] b_{ij}(o_{t+1}) \tag{3}$$

3. Termination:

$$P(O|\lambda) = \sum_{t=1}^{M} \sum_{k=1}^{K} \alpha_M(i,j,k) \tag{4}$$

3.2 Solution problem 2 for 2D

Parameters reestimation algorithm:

- Define $\xi(i,j,l)$ as the probability of being in state ij at time t and in state l at time $t+1$ for each k_{th} strem of data

$$\xi_t(i,j,l) = \frac{\alpha_t(i,j,k) a_{ijl} b_{ij}(o_{t+1}) \beta_{t+1}(i,j,k)}{P(O|\lambda)} = \tag{5}$$

$$\frac{\alpha_t(i,j,k) a_{ij} b_{ij}(o_{t+1}) \beta_{t+1}(i,j,k)}{\sum_{k=1}^{K} \sum_{l=1}^{N^2} \alpha_t(i,j,k) a_{ijl} b_{ij}(o_{t+1}) \beta_{t+1}(i,j,k)}$$

- Define $\gamma(i,j)$ as the probability of being in state i,j at time t, given observation sequence.

$$\gamma_t(i,j) = \sum_{l=1}^{N^2} \xi_t(i,j,l) \tag{6}$$

- $\sum_{t=1}^{T} \gamma_t(i,j)$ is the expected number of times state ij is visited
- $\sum_{t=1}^{T-1} \xi_t(i,j,l)$ is the expected number of transition from state ij to l

Update rules:

- $\bar{\pi_{ijk}}$ = expected frequency in state i,j at time $(t=1)$ $= \gamma_1(i,j)$

- $\bar{a}_{ij}$ = (expected number of transition from state i, j to state l) / (expected number of transitions from state i, j):

$$\bar{a}_{ijl} = \frac{\sum_t \xi_t(i, j, l)}{\sum_t \gamma_t(i, j)} \tag{7}$$

- $\bar{b}_{ij}(k)$ = (expected number of times in state j and oserving symbol k)/(expected number of times in state j):

$$\bar{b}_{ij}(k) = \frac{\sum_{t, o_t = k} \gamma_t(i, j)}{\sum_t \gamma_t(i, j)} \tag{8}$$

4 Experiments

4.1 Experiment 1

There was used *Amsterdam Library of Object Images* - the image database in this experiment. There are over one thousend colour images of small objects in this collection (Fig. 4). They systematically varied viewing angle, illumination angle, and illumination colour for each object, in order to capture the sensory variation in object recordings. They recorded over a hundred images of each object, yielding a total of 110,250 images for the collection [15, 16].

There has been selected fifty objects in testing the method. Three images for learning and three for testing has been chosen for each object. The 2D HMM has been implemented with parameters $N = 5, N^2 = 25, K = 25, M = 50$. Wavelet transform has been chosen as features extraction technique, and as wvelet function *db10*. Function selection has been made experimentally. Tab. 1 presents results of experiments.

Tab. 1: Comparison of recognition rate - pattern

Method	Recognition rate [%]
Eigenvector	94
1D HMM	90
2D HMM	92

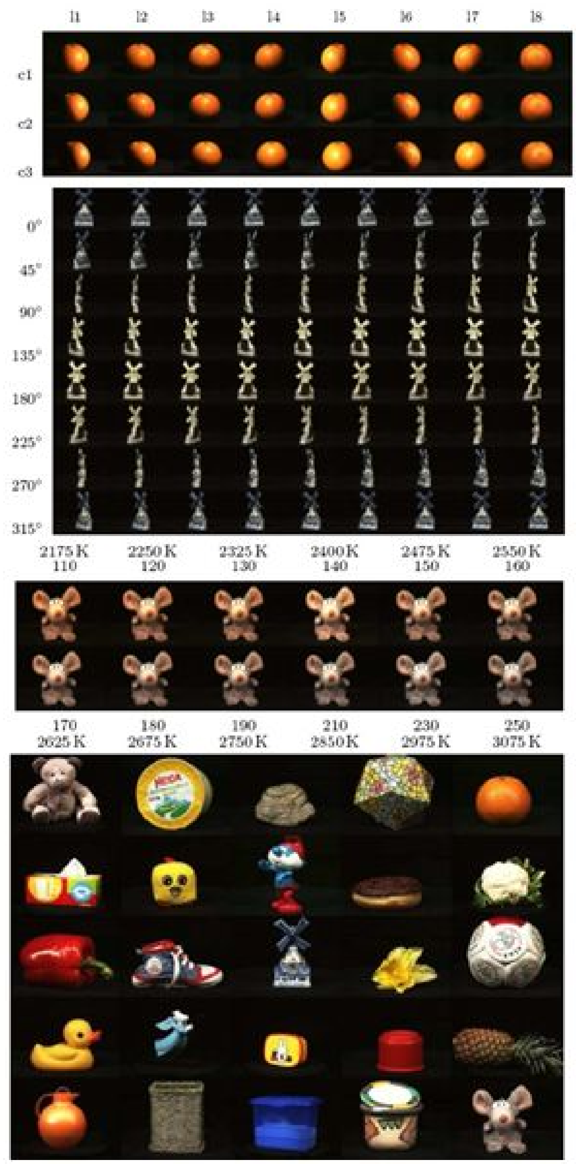

Fig. 4: Random representatives of the patterns in the ALOI dataset [17]

4.2 Experiment 2

The road signs image database German Traffic Sign Benchmark [17] was used in the experiment. There is over 1700 images of road signs in this database. The images show the signs in variable condition, lighting, rotation and size. Fig. 5 shows the example of traffic sign [18].

Fig. 5: Random representatives of the traffic sign in the GTSRB dataset [17]

Tab. 2: Comparison of recognition rate -road signs

Method	Recognition rate [%]
ESOM [19]	84
HMM [20]	49
1D HMM[our]	81
2D HMM[our]	83

We chose the 50 objects in order to verify the method, and for each object chose three images for learning and five for testing. The 2D HMM implemented with parameters $N = 4, N^2 = 16, K = 16, M = 25$. Wavelet transform was chosen as features extraction technigue, and *db10* as wavelet function. Function selection has been made experimentally. Tab. 2 presents the results of experiment.

5 Conclusion

New conception about two-dimensional hidden Markov models is presented. Presented method allows for faster image processing and recognition because they do not have to change the two-dimensional input data in the image form into a onedimensional data.
We show solutions of principle problems for ergodic 2D HMM, which may be applied for 2D data. Recognition rate of the method is 92% and 83%, which is better than 1D HMM. The advantage of this approach is that there is no need to convert the input two-dimensional image on a one-dimensional data, so we do not lose the information.
The obtained results are satisfactory in comparison to other method and proposed method may be the alternative solution to the others in unsupervised learning systems.

References

[1] L.R. Rabiner, A tutorial on hidden Markov models and selected application in speech recognition, *Proc. IEEE 77* pp. 257-285, 1989

[2] F. Samaria, S. Young, HMM-based Architecture for Face Identification, *Image and Vision Computing*, vol. 12, no. 8, pp. 537-583, 1994

[3] J.D. Hamilton, Analysis of time series subject to changes in regime, *Journal of econometrics*, vol. 45, pp. 39-70, 1990

[4] P. Baldi, Y. Chauvin, T. Hunkapiller, M.A. McClure Hidden Markov models of biological primary sequence information, *Proceedings National Academy of Science. U.S.A.*, vol. 91, pp. 1059-1063, 1994

[5] J. Bobulski, M. Kubanek, Person identification sys-

tem using sketch of the suspect, *Optica Applicata*, iss. 4, vol.42, pp. 865-873, 2012

[6] H. Valpola, Bayesian Ensemble Learning for Non-linear Factor Analysis, *Acta Polytechnica Scandinavica, Mathematics and Computing Series*, no. 108, Published by the Finnish Academies of Technology, 2000

[7] S. Eickeler, S Mller, G. Rigoll, High Performance Face Recognition Using Pseudo 2-D Hidden Markov Models, *European Control Conference*, http://citeseer.ist.psu.edu, 1999

[8] V.V. Bevilacqua, L. Cariello, G. Carro, D. Daleno, G. Mastronardi, A face recognition system based on Pseudo 2D HMM applied to neural network coefficients, *Soft Computing*, vol. 12, no. 7, pp. 615-621, 2008

[9] L. Yujian, An analytic solution for estimating two-dimensional hidden Markov models, *Applied Mathematics and Computation*, vol. 185, pp. 810-822, 2007

[10] J. Li, A Najmi, R.M. Gray, Image classification by a two-dimensional Hidden Markov model, *IEEE Transactions on Signal Processing*, vol. 48, pp. 517-533, 2000

[11] D. Joshi,J. Li, J.Z. Wang, A computationally Efficient Approach to the estimation of two- and three-dimensional hidden Markov models, *IEEE Transactions on Image Processing*, vol. 15, no. 7, pp. 1871-1886, 2006

[12] J. Bobulski, L. Adrjanowicz, Two-Dimensional Hidden Markov Models For Pattern Recognition, *Artificial Intelligence And Soft Computing, Lecture Notes in Artificial Intelligence*, Part I, 7894, pp. 515-523, 2013

[13] T. Kanungo, Hidden Markov Model Tutorial, http://www.kanungo.com, 1999

[14] G.D. Forney, The Viterbi Algorithm, *Procedings IEEE*, vol. 61, no. 3, pp. 268-278, 1973

[15] J.M. Geusebroek, G.J. Burghouts, A.W.M. Smeulders, The Amsterdam library of object images, *International Journal of Computer Vision*, vol. 61, no. 1, pp. 103-112, 2005

[16] Amsterdam Library of Object Images, http://aloi.science.uva.nl, 2013

[17] Database German Traffic Sign Benchmark, http://benchmark.ini.rub.de/Dataset/GTSRB-Final-Training-Images.zip

[18] J. Stallkamp, M. Schlipsing, J. Salmen, C. Igel, The German Traffic Sign Recognition Benchmark: A multi-class classification competition, *In Proceedings of the IEEE International Joint Conference on Neural Networks*, pp. 1453-1460, 2011

[19] Y.Y. Nguwi, S.Y. Cho, Emergent self-organizing feature map for recognizing road sign images, *Neural Computing and Application*, vol. 19, pp. 601-615, 2010

[20] J.C. Hsien, Y.S. Liou, S.Y. Chen, Road Sign Detection and Recognition Using Hidden Markov Model, *Asian Journal of Health and Information Sciences*, vol. 1 no. 1, pp. 85-100, 2006