Genetic Algorithm for parameters estimation and training within Hidden Markov Trees

YAMINA BORDJIBA[1] & HAYET MEROUANI[2]
Département d'Informatique
Université Badji Mokhtar
Faculté de l'Ingénieur
BP 12, 23000 Laboratoire LRI
Equipe Système de reconnaissance
des formes SRF
Annaba, ALGERIA

Abstract: - Texture analysis is an important task of various applications, such as medical imaging and robotic vision. Texture can be defined as a set of local neighbourhood proprieties of grey levels of an image region. One of the most important aspects of texture is the "scale". Psycho-visual studies indicate that the human visual system processes images in multiscale way.

A wide variety of texture analysis methods have been proposed [8], [11]. Wavelet transform intends to transform images into a multiscale representation, as the human system one [3]. Hidden Markov trees were proposed in 1998 for modelling the joint statistics of wavelet coefficients, i.e., interscale dependencies of wavelet coefficients [2]. HMTs were applied to image processing, such as image segmentation. EM algorithm is usually used for training, in this paper, a new training procedure is proposed, based on the genetic algorithms in order to improve models and increase their efficiency.

Keywords: - Wavelet Transform, Hidden Markov Trees, Genetic Algorithm, Texture detection, Image Analysis.

1 Introduction

Crouse and al. proposed a new statistical model in the wavelet domain, hidden Markov Tree (HMT), in 1998, which can models the joint statistics of wavelet coefficients, i.e., interscale dependencies of wavelet coefficients [2]. HMTs are particular hidden Markov Models (HMMs), structured in trees instead of chains, connecting states variables of wavelet vertically across scale, however, three problems have to be resolved, hidden state estimation, likelihood determination and training. Maximum A Posteriori (MAP) algorithm is used to resolve the first problem. As there is no formula to compute the likelihood, so the upward-downward algorithm inspired from the forward-backward algorithm is used in HMMs. Model parameters estimation and training are traditionally achieved with Expectation Maximisation (EM) algorithm. An accurate estimation of HMT parameters is essential for the implementation. However, EM algorithm converges generally to a local optimum. Several no stochastic and stochastic versions of this algorithm were proposed, but none of them have resulted in satisfactory version, hence our proposition to use genetic algorithm, which allows the exploration of large research space. Genetic Algorithm (GA) computes simultaneously several solutions instead of a single one, exactly in the same way as in the EM algorithm. At each step of the genetic algorithm, combining individuals of current population, applying genetic operators, selection, crossover and mutation creates a new population.

In the second section of this paper, we define Hidden Markov Model, then we introduce Hidden Markov Tree and we discuss their using problems as well as the adopted solutions. In the last section, the EM algorithm is presented as genetic algorithm, that what we proposed for HMT training.

2 HMT via HMM

HMTs are particular HMMs that are structured in trees instead of chains. For this reason, we are leaded up to present HMM first and then HMT.

2.1 Hidden Markov Model

HMMs were developed in 1960s [5] and have been a successful in the domain of speech recognition, thanks to their context integration capacity and noise absorption and since then many varieties of them have been applied in other domains such as writing recognition [1].

An HMM is a finite set of states transitions among the states are governed by "transition probabilities", and a set of observations. So, HMM can be defined by two random variable suites, one hidden and the other observable [1].

An HMM is noted by $\lambda=(A,B,\pi)$ where:

M is the number of states of the model.

N is the number of observations symbols.

A is a state transition probabilities matrix, $A=\{a_{i,j}\}$,

$a_{i,j} = p(q_{t+1}=j / q_t=i)$ where :

$1 \leq i, j \leq M$.

with q_t the current state

and q_{t+1} the next state.

B is a symbol of probabilities matrix observation,

$B=\{b_j(k)\}$, $b_j(k) = p(o_t=v_k / q_t=j)$ where :

o_t the current observation.

and q_t the current state.

v_k a symbol of observation.

$1 \leq j \leq M$.

$1 \leq k \leq N$.

π represent the initial state of distribution,

$\pi = (\pi_i)$, $\pi_i = p(q_1=i)$ where

q_1 is the initial state.

$1 \leq i \leq M$.

Given an HMM, there are three problems :

- ***Evaluation problem***

Given an HMM λ and a sequence of observations *O*, what is the probability that the observations are generated by the model, i.e., $p(O/\lambda)$?

To solve this problem, a solution was proposed by using forward-backward algorithm, which allows to estimate this probability [1].

- ***Training problem***

Given an HMM λ and a sequence of observations *O*, how should we adjust the model parameters in order to maximise $p(O/\lambda)$?

For this problem, the solution is the use of Expectation Maximisation algorithm "EM"[1].

- ***Decoding problem***

Given an HMM λ and a sequence of observations *O*, what is the what is the most likely state sequence in the model that produced the observations?

To solve this problem, the basic solution generally used is the Viterbi algorithm[1].

HMMs have earned their success in form recognition, and more particularly in speech recognition, however, their application to texture analysis is a very difficult task, as the space is in two-dimensional. Even, if in planar HMM has been proposed, it can not be used for texture analysis, as it is necessary to analyse the neighbourhood of pixels.

For that reason, Crouse and al. [2] have proposed a new statistical model HMT, which is based on two theories, HMM and wavelet. In fact, the HMTs are HMMs that are structured in trees instead of chains, connecting states variables of wavelet coefficients vertically across the scale.

2.2 Hidden Markov Tree

wavelet transform intend to transform the image into a multiscale representation with both spatial and frequency characteristics which allow an effective multiscale analysis of the image. The HMT can model the joint statistics of the wavelet transform by capturing the interscale dependencies of wavelet coefficients[4].
Known that observable states are wavelet coefficients $w_{j,i}$ and hidden states are $S_{j,i}$, where J is the number of scales of the wavelet transform.
An M-state HMT is defined by :

- The distribution for the root node S_1 : $\pi = (\pi_m)$ with $m \in \{1,\ldots,M\}$.
- The transition probabilities $A= \{ a^{mn}_{\rho(i),i} \}$ where
  i : varying from 1 to the total wavelet number
  $m,n = 1\ldots M$.
  $\rho(i)$: is the parent of the node i.
  $a^{mn}_{\rho(i),i} = p(S_i = n / S_{\rho(i)} = m)$
- The mean $\mu_{j,m}$ and the variance $\gamma^2_{j,m}$ of wavelet coefficient $W_{j,i}$ where $j=1\ldots J$ et $m=1\ldots M$.

The model is noted by $\lambda= (\pi, A, \mu_{j,m}, \gamma^2_{j,m})$ $j=1\ldots J$ and $m=1\ldots M$.

Fig.1

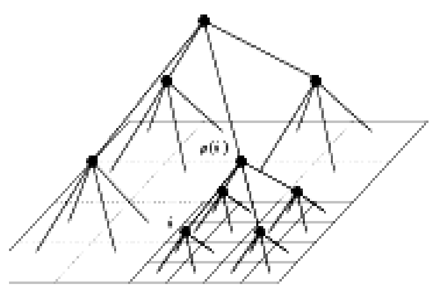

Fig.2

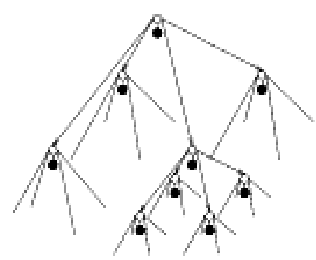

Fig.1 represents an image decomposition by wavelet transform which is structured in a tree [7], and the correspondent HMT is represented by the Figure (2). Black nodes denote wavelet coefficients and white nodes denote the correspondent hidden states and the links denote the dependencies between states (the parents nodes and their sons).

In [5], authors propose several statistical models for the signal modelisation, and they use HMTs for texture classification and segmentation via the model HMT-3S. Simulation results have shown that this model has improve the classification and the texture segmentation accuracy. The model HMT-3S groups the three wavelet sub-bands into one quad-tree structure. Therefore, this model is similar to the original HMT expect for the number of the wavelet coefficient in each node, i.e., hidden states in a node.

In [12], authors propose an algorithm for image segmentation using HMTs, after the application of the wavelet transform on the image, they use HMTs to classify all dyadic squares at all scales. Segmentation algorithm is as follows:

1. train wavelet domain HMT for each texture, using EM algorithm.
2. compute the likelihood of each dyadic square at each scale.

3. model the dependencies between different dyadic squares using a labelling graph.

As with HMMs, given an HMT, there are also three problems, hidden states estimation, likelihood determination and training. To resolve them, we use MAP algorithm for the first one. As there is no formula to compute likelihood, the problem is solved using upward-downward algorithm inspired from the forward-backward algorithm used in HMM. Parameters model estimation and training are traditionally realised by EM algorithm.

3 HIDDEN MARKOV TREE TRAINING

In this paragraph, we are interested by the training problem. For that, an upward-downward algorithm is presented firstly. EM algorithm, realise the step of training.

3.1 Upward-Downward Algorithm

It consists on two steps, an ascendant one, called upward and a descendant one, called downward.

Upward step : consists of computing the joint probability of each sub-tree of the entire tree and starts from terminal nodes.

Downward step : consists of computing the joint probability of the entire tree and starts from the root of the tree.

3.2 Training by EM algorithm

An accurate estimation of the HMT parameters is essential for practical applications, it can be achieved using EM algorithm. Crouse proposed the first version adapted to the HMT model[2], the basic idea is to construct, at each iteration, a refined model comparing to the precedent one, starting by a model that is estimated randomly.

HMT training using EM algorithm amount to fit an M-state model λ to the observed J-scale tree-structured, i.e., W. The algorithm can be described as follows :

Step 1 : initialisation
Set an initial model λ^0, and iteration counter i=0.

Step 2 : E step
Calculate the probability $p(S,W/\lambda^i)$ for hidden state variables.

Step 3 : M step
Calculate λ^{i+1} parameters (known the probability for hidden states) λ model parameters can be updated.

Step 4 : iteration
set i=i+1. if it converges, then stop; else return to step 2.

3.3 Training by Stochastic Versions of EM Algorithm

EM algorithm allows to estimate and train HMT model. However, it is sensitive to the initial parameters settings and often it converges to a local optimum. Several authors have proposed various non stochastic improvements of the EM algorithm none of these improvements resulted in a completely satisfactory version of EM. To overcome to EM's limitations, three stochastic versions was proposed: SEM, SAEM and MCEM.

SEM is a variant of EM algorithm, its basic idea is to incorporate a stochastic step between E and M steps, SEM can be seen as a stochastic perturbations of the EM algorithm. In case of small set of data, the variance of SEM is large, for this reason, SAEM algorithm was proposed, it is a modification of the SEM algorithm. SAEM is going from pure SEM at beginning towards pure EM at the end. Theoriclly, it was shown that, almost certainly, SAEM converges to a local maximum likelihood. However, a slow rate of convergence is necessary for good performance. So, another stochastic algorithm was proposed, MCEM algorithm for models where expectation is too complex to allow a direct maximization. Under suitable technical conditions, MCEM algorithm converges to a stationary point of likelihood.

3.4 Training by Genetic Algorithm

To overcome the above mentioned limitations of EM algorithm, we propose a new method that uses genetic algorithm for parameters model estimation and training. In this last section, we search for the optimal model producing observations, for this reason, the use of genetic algorithm has been used, known for its power in the domain of optimisation.

GA was developed in 70s by Holland [6]; the idea is to reproduce the natural evolution of individuals, generation after generation, respecting inheritance phenomena. Starting from an initial generation (called also population), GA generates several new populations, applying genetic operators. The selection allows choosing parents for the reproduction using adaptation function. The crossover consists of obtaining of new individuals, combining a pair of parent's individuals, and the mutation that amounts to change one or many symbols in the same individual. The aim of this last is to diversify individuals of the population [10]. The algorithm can be described as follows:

Initialisation

Random generation of the initial population

Evaluation of adaptation function

Calculate the likelihood of each individual of the actual population, using upward-downward algorithm.

Stop test

If the maximum number of iteration (fixed before) is reach or an optimal model is found then go to ending.

Selection

Using values calculated in evaluation phase, best individuals are chosen to be used like parents in crossover; new individuals created by crossover operator will replace rest of population.

Crossover

We propose to use the crossover at one point.

Mutation

We propose the following mutation operator:

- A number P of positions to be modified is randomly chosen.
- P positions (integer) are chosen randomly.
- for each position i, from the P positions, a real number between 0 and 1 is so randomly chosen.

Normalization

This operator must be applied on each individual modified in the last phase.

Looping

Return to evaluation phase.

Ending

Decode the best solution in the actual population.

Individuals of our genetic algorithm are HMts, they are coded by concatenating all HMT's parameters and converting all matrix in vectors [9]. So, individuals are vectors of real numbers. The initial population is randomly generated, then, we apply genetic operators. Before the use of these operators, we must fix crossover and mutation points, and the adaptation function for the selection. For our application, the fitness will be the likelihood of the HMT model.

The use of normalisation operators is own to HMT training problem, it is used to verify solution coherence, all matrix in HMT's parameters must be stochastic.

4 Conclusion and perspectives

In this paper, we have presented HMT combined to GA to their use in texture analysis. Since their proposition, they were used in several applications such segmentation. However, these applications have proved that their training, using EM algorithm, was enough heavy and converge generally to a local optimum, that what we motivated to use GA in training, for their success in optimisation. Our perspectives are to apply this hybrid method to radiographic images, and more precisely in texture segmentation in order to proof its efficiency and to show its improvement compared with EM algorithm use.

References :

[1] A.Belaïd, G.Saon , Utilisation des processus Markoviens en reconnaissance de l'écriture. Traitement du signal, Vol. 14 , n° 2, 1997.
[2] M.S.Crouse, R. D. Nowak, and R. G. Baraniuk, Wavelet-based statistical signal processing using hidden Markov models. IEEE Trans. Signal Proc., 46(4), 1998, pp. 886–902.
[3] J.Durant, P.Gonçalvès, Statistical Inference for Hidden Markov Tree Models and Application to Wavelet Trees, Rapport Technique RR-4248, INRIA, Sep.2001.
[4] G.Fan, X. Xia, Maximum Likelihood Texture Analysis and Classification Using Wavelet-Domain Hidden Markov Models. In proc.of the 34th Asimolar conference on signals, systems and computers, Pacific Grove, CA, Oct. 29 – Nov. 1, 2000.
[5] G.Fan, wavelet domain statistical image modelling and processing. Ph.D. dissertation, University of Delaware, May 22, 2001.
[6] J.H.Holland, Adaptation in natural and artificial systems. Ann Arbor, University of Michigan Press, 1975.
[7] J.K.Romberg, H.Choi and R.G.Baraniuk, Bayesian tree structured image modelling using Wavelet-Domain Hidden Markov Models. Proc. of SPIE, Denver, Vol. 3816, July 1999, pp. 306-320.
[8]T.R.Reed, and J.M.H.Buf, Image Understanding, CVGIP, 57, 1993, pp.359-372.
[9] M.Slimane, T.Brouard, G.Venturini and J.P.A Debeauville, Apprentissage non-supervisé d'images par hybridation génétique d'une chaîne de Markov cachée, Traitement du Signal, Vol. 16, N° 6, 1999, pp. 461-475.
[10]A.Soukhal, W.Duque, N.Sbiai and P.Martineau, Application des algorithmes génétiques aux Flowshop hybrides. Conférence Internationale en Informatique, CII'99, 21-23 Annaba, Algérie, Nov. 1999.
[11] Tuceryan, M. Et Jain, A. K., Handbook of Pattern Recognition and Computer Vision, World scientific, Singapore, 1993.
[12] Zhen Ye, Image segmentation using Wavelet-Domain Hidden Markov Models. Technical report. Dec. 2000. Kent State University.