COMBINING PARTICLE SWARM OPTIMISATION WITH GENETIC ALGORITHM FOR CONTEXTUAL ANALYSIS OF MEDICAL IMAGES

Jonathan Goh[1], Lilian Tang[1], Lutfiah Al turk[2] and Yaochu Jin[1]

[1]*Department of Computing, University of Surrey, Guildford, GU2 7XH, Surrey, U.K.*
[2]*Department of Statistics, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia*

Abstract: Micro aneurysms are one of the first visible clinical signs of diabetic retinopathy and their detection can help diagnose the progression of the disease. In this paper, we propose to use a hybrid evolutionary algorithm to evolve the structure and parameters of a Hidden Markov Model to obtain an optimised model that best represents the different contexts of micro aneurysms sub images. This technique not only identifies the optimal number of states, but also determines the topology of the Hidden Markov Model, along with the initial model parameters. We also make a comparison between evolutionary algorithms to determine the best method to obtain an optimised model.

1 INTRODUCTION

Micro aneurysms are one of the first visible signs of Diabetic Retinopathy (DR) and it is known that quantities of this clinical sign can help diagnose the progression of the disease. Micro aneurysms are swelling of the capillaries that are caused by the weakening of the vessel walls due to high sugar levels in diabetes and eventually leak to produce exudates. In retina images, micro aneurysms appear as small reddish dots with similar intensity as haemorrhages and blood vessels. This particular sign is an important early indicator of the disease and can contribute to helping ophthalmologists identify effective treatment for the patient at an early stage.

However, an accurate detection of micro aneurysms is a challenge task. One of the main obstacles is the variability in the retinal image, depending on factors such as degree of pigmentation of epithelium and choroid in the eye, size of pupil, illumination, disease, imaging settings (which can vary even with same equipment), patients' ethnic origin, and other variants. These factors affect the appearance of micro aneurysms. They tend to appear among other visual features and the difference between a micro aneurysm and its surroundings can be very subtle.

Standard image processing and classification techniques alone are not able to deal with the ambiguity in micro aneurysm detection. They are often mistaken as other similar visual content in retinal images such as the fine ends of the blood vessels or noise. In the work reported by Niemeijer et al. (2005) and Sinthanayothin et al. (2002) image processing techniques were first adopted to extract useful features followed by recognition through a classifier. However, the single classifier used is unable to ensure scalability. Walter et al. (2000) developed a technique that requires the blood vessels to be removed prior to micro aneurysm detection and as a result, true micro aneurysms near or on the blood vessels are removed as well. This suggests that the recognition procedure of this clinical sign cannot be treated in isolation. Instead, an integrated approach that dynamically combines detection evidence from various processing stages, and especially a contextual environment each time the clinical sign may appear should be constructed. In our research, we developed multiple classifiers together with a contextual reasoning model to address the scalability and ambiguity. In this paper we mainly discuss the contextual model.

Hidden Markov Models (HMMs) is a statistical modelling tool for information extraction. While HMMs have been successful in many applications

such as speech recognition (Morizana et al., 2009 and Lu et al., 2009) DNA sequencing (Won et al. 2006) and handwriting recognition (Parui et al. 2008) very little work has been carried out to statistically model and understand the context in images. In speech recognition, HMMs can determine the statistical variations of utterance from occurrence to occurrence. However, a few outstanding issues remain. Firstly, how to determine the topology of the HMM and secondly, what is the optimised model parameters for accurate representation of the training data? Lastly, the training of HMM is computationally intensive and there is no known method that can guarantee an optimised model.

Optimising an HMM is usually done through the refinement of the HMM after each training. Refinement can include changing the number of states, the initial distribution states and the transition probabilities before re-training the HMM and testing it for its accuracy. The most popular training algorithm for HMM is the Baum-Welch algorithm; however, this algorithm is a hill climbing algorithm and heavily depends on the initial estimates. It is also known that bad estimates for this algorithm usually lead to a sub-par HMM. Hence, the motivation behind this work is to obtain an optimised HMM based on the initial parameters used to train a HMM.

Evolutionary algorithms (EAs) have shown to be powerful in solving difficult optimisation problems. Most of the published work such as Won et al. (2006), Kwong et al. (2001), Bhuriyakorn et al. (2008) and Xiao et al. (2007) uses EAs to optimise HMM using a combination of Genetic Algorithms (GAs) and the Baum-Welch Algorithm (BW). However, these techniques only determine the optimal number of states and improves BW generalisation. The main idea of this work is to optimise the topology of the HMM while adapting the parameters over the evolutionary process for an optimised model.

Memetic Algorithms (MAs) are a class of hybrid algorithms that combine a population-based global heuristic search strategy with a local refinement (Ong et al. 2010). MAs have been reported to be successful in multiple domains such as scheduling (Lim et al. 2005), machine learning (Liu et al., 2007 and Abbass, 2002) and even aerodynamic design optimization (Ong et al. 2003).

Our previous work (Goh et al. 2010) has demonstrated the effectiveness of HMMs in the detection of micro aneurysms as a contextual analysis model. In this paper, we extend our previous work by using a combination of a Genetic Algorithm and Particle Swarm Optimisation (*referred as Memetic Algorithm from here on*) to optimise the structure of the HMM. In Section 2, we give a brief description of the Memetic Algorithm and HMMs. The technique used for optimising the HMM is presented in Section 3. Section 4 describes the experiments and we summarise our work in section 5.

2 EVOLUTIONARY ALGORITHMS & HIDDEN MARKOV MODELS

Memetic algorithms use different search techniques in a combined approach and maintain a population of solutions. The main difference is that for every solution, a local-improver will be used to further enhance the solution.

A Genetic Algorithm is used to perform the global search, as it is a population-based stochastic search method whereas for the local search, we use Particle Swarm Optimisation (PSO). At each generation of the GA, a new set of solutions is created by a process of selecting individuals according to their strengths (fitness) in the problem domain and genetically modifying them to produce offspring. This process leads to the generation of a new population of individuals that are better suited for the problem than the individuals that they are created from, eventually reaching an optimal solution.

For each solution, PSO will be carried out to further optimise the solution. PSO functions by propelling the particle (individual solution) through the search space with a velocity that is dynamically modified based on its own strength and the strength of other particles in the swarm.

Ideally, after the termination criteria have been met, the final population would consist only of the best individuals which would be decoded as the optimised set of solutions.

In our work, each solution would be encoded into a chromosome which represents the HMM structure. Typically, a HMM is characterised by:

a) Number of states, M
b) Transition probability distribution matrix A. A=$\{a_{ij}\}$, where a_{ij} is the transition probability of the Markov chain transiting from state i to state j.
c) Observation sequence, O.
d) Initial state distribution, π.

Hence, the HMM is represented by: $\lambda = (A, O, \pi)$. In order for the HMM to represent the image effectively, we need to decide upon the topology of the HMM, the number of states of the model and the transitions that are allowed between states.

Training of the HMM can be carried out using the BW algorithm which is an expectation maximisation algorithm that adjusts the model parameters to locally maximise the likelihood of the training data based on an initial estimate of the parameters.

Recognition of the image is performed using the Viterbi algorithm which finds the most likely state sequence given the HMM model, λ and a sequence of observations.

The percentage accuracy is calculated as the total number of correctly predicted images over the total number of images.

3 HMM EVOLUTION

In order for a HMM to effectively represent the training data, the number of states and the structure of the connecting states are crucial.

In the following sections, we demonstrate the use of the memetic algorithm to optimise HMMs using sub-images of micro aneurysms (MA), background (BG) and blood vessels (BV) as the training data. A GA will be used to evolve the structure of the HMM while PSO will be used to optimise the parameters for the HMM as detailed in the pseudo code in Figure 1. By performing a hybrid search using the memetic algorithm, a balance between exploration and exploitation can be achieved. This evidently not only automates the discovery of HMM structures along with the initial model parameters, the resulting model can also attain a better accuracy while avoiding overfitting, as we will discuss later on in the section.

```
Initialise Population
While iteration < Max_Generation
    SelCh = Selection(population);
    SelCh = CrossOver(SelCh);
    FitterSolutions = bestSolutions(SelCh);
    For all_of_FitterSolutions
    New_solution = PSO(FitterSolutions)
        If New_solution > SelCh
        SelCh = New_solution
    endIf
        endFor
    population = recombination(SelCh);
endWhile
```

Figure 1: Pseudo code of Algorithm.

3.1 Feature Extraction for HMM

The training data used for this research are 15 by 15 pixel images which are the output from the ensembles in our earlier work (Goh et al. 2010), which comprise of micro aneurysms (MA), background (BG) and blood vessels (BV).

Each sub-image is divided into nine 5x5 pixel smaller sub-images as seen in Figure 2, which are used as observation sequences for the HMM.
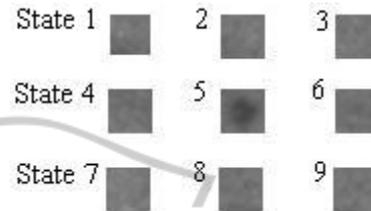

Figure 2: States of Sub-Image.

The *Discrete Cosine Transform* (DCT) is performed to obtain the features for each of the 5x5 pixel sub-image. The DCT is used as it can represent an image in terms of sum of sinusoids of varying magnitude and frequencies, thus obtaining the most important information in terms of just a few coefficients. Once the DCT has been applied for each observation, the result from the DCT process for each state is reshaped into a 25x1 column and used as part of a sequence for inputting into the HMM.

3.2 Global Search - GA

For optimisation, the solution has to be encoded into a chromosome for evolution. In this work, since HMM uses real-valued numbers, a real-valued string was used as the chromosome in the GA. The chromosome consists of the following information:

1. Number of states
2. Type of states as seen in Figure 3
3. Transition probabilities

3.2.1 Initial Population

The initial population was generated randomly. For each candidate solution, a number of states, which is an integer between 4 and 11, was randomly generated. This is based on Bakis' (1976) assumption that the number of states is usually identical to the number of the observed sequences. In this work, nine observation sequences are used to represent the various sub-images, thus the minimum number of states is set to 4 and the maximum

number of states to 9. With the initial number of states, the transition between states can be set.

For each state, there are a few different kinds of transitions that can be assigned to them as listed in Figure 3 and they are randomly assigned to each state. Initial state transition probabilities are also randomly assigned between the initiating states and the transiting states.
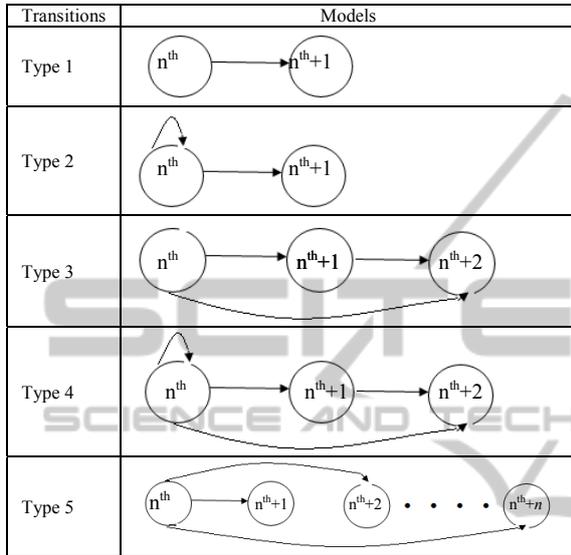
Transitions	Models
Type 1	
Type 2	
Type 3	
Type 4	
Type 5	

Figure 3: Transition types.

3.2.2 Fitness Evaluation

In order to measure the generalisation capability of the HMM for recognising micro aneurysm sub-images, we use a fitness evaluation mechanism to gauge the confidence level of each solution. Initially, we used the average maximum likelihood that is calculated by the BW algorithm to measure the fitness used in selecting fitter individuals from the population. The average maximum likelihood p_n of the HMM, λ, that generates the observation sequence $O_1, O_2 \ldots O_n$ is calculated using the following equation:

$$p_n = \left(\sum_{n=i}^{T} p(O_n \mid \lambda) \right) / T$$

where T is the number of observation sequences for training.

However, our analyses showed that generalising the average maximum likelihood does not necessarily produce a better accuracy due to over-fitting of the training data. Hence, in this work, we use the accuracy obtained from the last re-estimation of the BW algorithm as the fitness value.

3.2.3 Selection

Selection is the phase used to determine which parents to choose for reproduction. In this work, we chose to use the Roulette Wheel Selection (RWS). The advantage of RWS is that they may allow weaker individuals still to be selected for reproduction as they may have important components that may be useful during the recombination process. The parameter used in selection is set at 0.8, that is to say, 80% of the population are selected for crossover and mutation. However, local search using the PSO is applicable only to the top 20% of the best individuals after selection.

3.2.4 Crossover

This operation represents the major driving force in the canonical GA for optimizing the structure of the HMM. In crossover, we need to decide on a crossover point to swap parts of chromosome of the parents to produce offspring. In this work, we adopted the 1-point crossover.

If both parents have the same number of states, the creation of offspring is straightforward. However, if the two parents have different number of states, there must be a decision on how many states the offspring will have. For simplicity, we assume that the offspring shall have the average number of states between the two parents. To make up for the additional state, the offspring will inherit the additional state for the parent as illustrated in Figure 4.
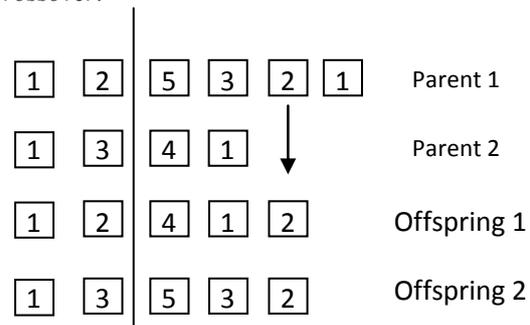
Crossover:

Figure 4: Crossover Operation.

3.3 Local Search - PSO

As the BW algorithm is very sensitive to the initial model parameter, in order to exploit the local search

Table 1: Comparison between different Evolutionary Algorithms.

Pop	Gen	Average Maximum Likelihood/Accuracy					
		Memetic Trained HMM (M-HMM)			GA Trained HMM (GA-HMM)		
		MA Models	BV Models	BG Models	MA Models	BV Models	BG Models
30	30	-8.1209/ 96.41%	-8.3076/93.25	-8.0949/91.04%	-8.1253/96. 19%	-8.294/92.64	-8.1241/90.49%
30	60	-8.1430/ 96.86%	-8.2997/93.36	-8.1038/91.04%	-8.1109/ 96.19%	-8.304/92.33	-8.1297/91.22%
50	30	-8.1273/97.04%	**-8.3076/94.79**	-8.07650/91.41	-8.1256/ 93.95%	-8.3035/93.25	-8.0969/91.22
50	60	**-8.1394/97.09%**	-8.3132/92.64	**-8.0783/91.77%**	-8.1366/96.86%	-8.298/94.17	-8.0978/91.6%

Table 2: Comparisons among various methods.

Models	Average Maximum Likelihood				
	Optimised M-HMM (9)	Optimised GA-HMM (9)	BW Trained HMM		
			7 States	8 States	9 States
MA	-8.1394	-8.1366	-8.215	-8.209	-8.150
BV	-8.3076	-8.2928	-8.378	-8.342	-8.328
BG	-8.0783	-8. 0978	-8.274	-8.252	-8.186

region for better solutions, we apply the PSO to the top few individuals obtained after selection.

The PSO starts from individual chromosomes resulting from the GA search and the its goal is to find optimised transition probabilities for potentially good solutions. For the states which were inherited during evolution, no new transition probabilities are generated. For the newly generated states, the transition probabilities are randomly generated to allow the PSO to search the locally around the solutions.

For PSO, we use a swarm size of 10 particles for 30 iterations.

4 EXPERIMENTAL RESULTS

4.1 Data Set

The 15 by 15 training samples used to train Hidden Markov Models are obtained from 100 retina images of various sources including the Optimal Detection and Decision-Support Diagnosis of Diabetic Retinopathy database.

4.2 Experiment Setup

700 background (BG) sub-images, 700 micro aneurysms (MA) sub-images and 700 blood vessel (BV) sub-images are used to train the different HMMs. In order to test the accuracy of the models, we have a test set that contains the 3 categories with each one consisting of 500 sub-images.

4.3 Experiment Results

The Memetic-HMM (M-HMM) algorithm was run according to the parameters setup given in Table 1 for optimising the various models and their average maximum likelihood along with their accuracy are listed after the relevant generations were reached.

Considering the results listed in Table 1 along with the algorithm parameters, we compare these results with those obtained by using a GA only, termed HMM (GA-HMM). The GA-HMM follows the same steps described in Section 3.2.1 – 3.2.3, the only difference is that in the latter GA handles the mutation of the Transition Probabilities instead of the PSO.

Our results show that although the Average Maximum Likelihood is higher, it does not necessarily mean a better accuracy as we can see that the MA models labelled in grey has a lower average maximum likelihood compared to the GA HMM but a higher accuracy.

This suggests that by using the memetic algorithms, the parameters for each solution are adaptive over the evolutionary process allowing for the optimised structure of the HMM while adapting the transition probabilities for the optimised structure. It also suggests that this technique reduces the risk of over-fitting the training data since the fitness evaluation is the accuracy rather than continuous training for the highest average maximum likelihood that may eventually causes overfitting.

For the rest of the models, the memetic algorithm is able to obtain both better accuracy and generalisation compared to the GA only approach. Naturally, for each model, we use the model with the

highest accuracy. The performance listed in Table 2, indicates that the optimal number of states found by the both evolved HMMs are identical to a manually trained HMM. It also indicates that they are far more optimised than a manually hand designed HMM using the BW algorithm.

4.4 Experimental Performance

While the difference between the M-HMM and the GA-HMM is not significantly large, comparing the number of generations for the population based search, using memetic algorithms to evolve the HMM results in a faster convergence to an optimal solution as illustrated in Table 3.

Table 3: Convergence Times.

Model (Pop/Gen)	Convergence Generation	
	M-HMM	GA-HMM
MA (50/60)	4th	34th
BV (50/30)	13th	27th
BG (50/60)	15th	43rd

5 CONCLUSIONS

In summary, a novel way to represent images using a fully automated structure discovery technique involving Memetic Algorithms and HMM was presented in this paper. A comparison was made between various methods and the experimental results have shown that M-HMM is capable of searching for a more optimal structure than that resulting from either the GA only approach or the BW Algorithm.

By using evolutionary algorithms to evolve the HMM, we can not only find the optimal number of states to represent the image, but also manage to optimise the initial transition probabilities for a better trained model as indicated by its average maximum likelihood. Although the recognition rate of the M-HMM is just slightly better than the GA-HMM, the former converged quicker to optimal solutions suggesting that memetic algorithms can be applied to situations where time is of an essence. These results demonstrate that the EA evolved HMMs are capable of context reasoning for detecting micro aneurysms and thus facilitate finer analysis during clinical sign detection on retina images.

REFERENCES

Abbass, H. A., 2002. "An evolutionary artificial neural networks approach for breast cancer diagnosis", *Artificial Intelligence in Medicine*, Vol. 25.

Bakis, R., 1976. Continuous speech word recognition via centisecond acoustic states, *Proceedings ASA Meeting*, Washington,DC.

Bhuriyakorn, P., Punyabukkana, P., Suchato, A., 2008. "A Genetic Algorithm-aided Hidden Markov Model Topology Estimation for Phoneme Recognition of Thai Continuous Speech" *Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing.*

Goh, J., Tang, H. L., Al turk, L., Vrikki, C., Saleh, G., 2010. "Detecting Micro aneurysms using Multiple Classifiers and Hidden Markov Model", 3rd International Conference on Health Informatics, Valencia, Spain.

Kwong, S., Chan, C. W., Man, K. F., Tang, K. S., 2001. "Optimization of HMM topology and its model parameters by genetic algorithms", *Pattern Recognition,* 34:509-522/

Lim, M. H., Xu, Y. L., 2005. "Application of Hybird Genetic Algorithm in Supply Chain Management," Special Issue on Multi Objective Evolution: Theory and Applications, Intternational Journal of Computers, Systems and Signals, Vol. 6.

Liu, B., Wang, L., Jin, Y., Huang, D., 2007. "Designing Neural Networks using PSO-Based Memetic Algorithm", Advances in Neural Networks, Vol. 4493.

Lu, G., Jiang, D., Zhao, R., 2007. "Single Stream DBN Model Based Triphone for continuous speech recognition", *Proceedings of the 9th IEEE International Symposium on Multimedia Workshop*.

Meindert, N., Ginneken, B., Stal, J., Suttorp-Schulten, M., Abramoff, M., 2005. "Automatic Detection of Red Lesions in Digital Color Fundus Photograph", *IEEE Transaction on Medical Imaging*, Vol. 25(5).

Morizane, K., Nakamura, K., Toda, T., Saruwatari, H., Shikano, K., 2009. "Emphasized Speech Synthesis Based on Hidden Markov Models", 2009 Oriental COCOSDA International Conference on Speech Database and Assessments.

Ong, Y. S., Lim, M. H., Chen, X. S., 2010. "Research Frontier: Memetic Computation – Past, Present & Future", *IEEE Computational Intelligence Magazine*, In Press.

Ong, Y. S., Nair, P. B., Keane, A. J., 2003. "Evolutionary Optimisation of Computationally Expensive Problems via Surrogate Modelling", *American Institute of Aeronautics and Astronautics Journal*, Vol. 41.

Parui, S. K. ., Guin, K., Bhattacharyam, U., Chaudhuri, B. B, 2008. "Online Handwritten Bangla Character Recognition Using HMM", *IEEE Transaction* 2008.

Sinthanayothin, C., Boyce, J. F., Williamsom, T. H., Cool, H.L., Mensah, E., Lai, S., Usher, D., 2002. "Automated Detection of Diabetic Retinopathy on Digital Fundus Images", *Diabetic Medicine*, Vol. 19, ppl105-112.

Walter, T., Klein, J. C., 2000. "Automatic Detection of Micro aneurysms in Color Fundus Images of the Human Retina by means of the Bounding Box Closing", *Proceedings of the 3rd International Symposium on Medical Data Analysis*, Rome, Italy.

Won, K. J., Prugel-Bennet, A., Krogh, A., 2006. "Evolving the Structure of Hidden Markob Models", *IEEE Transaction on Evoutionary Computation*, Vol. 10 (1).

Xiao, J. Y., Zou, L., Li, C., 2007. "Optimization of hidden Markov model by a genetic algorithm for web information extraction", *Proceedings of the 2007 International Conference on Intelligent Systems and Knowledge Engineering*, Chengdu, pp. 153-158.