

Optimization of HMM by the Tabu Search Algorithm

TSONG-YI CHEN, XIAO-DAN MEI*, JENG-SHYANG PAN
AND SHENG-HE SUN*

*Department of Electronic Engineering
National Kaohsiung University of Applied Sciences
Kaohsiung, 807 Taiwan*

**Department of Automatic Test, Measurement and Control
Harbin Institute of Technology
Harbin, 150001 China*

In this paper, a simple version of the tabu search algorithm is employed to train a Hidden Markov Model (HMM) to search out the optimal parameter structure of HMM for automatic speech recognition. The proposed TS-HMM training provides a mechanism that allows the search process to escape from a local optimum and obtain a near global optimum. Experimental results show that the TS-HMM training has a higher probability of finding the optimal model parameters than traditional algorithms do.

Keywords: tabu search, hidden Markov model, forward-backward algorithm, speech recognition, global optimum

1. INTRODUCTION

HMM is a highly robust statistical method widely used for automatic speech and speaker recognition. It is a powerful algorithm used to estimate the model parameters, and it can achieve a high level of performance [1-10]. HMM has been successfully combined with bounded state durations for use in isolated word recognition [2]. The weighted HMM and subspace projection algorithm have been applied to robust HMM-based speech recognition [3]. HMM has also been applied to speaker adaptation [4-6] and noisy speech recognition [7, 8]. Once a structure of the HMM model is given, the model parameters can be obtained automatically by feeding training data. The HMM model parameters play important roles in a HMM based speech recognizer because they can characterize the behavior of the speech segments and directly affect the system recognition accuracy.

Many heuristic algorithms have been developed to optimize the model parameters in order to better describe the trained observation sequences. These include the forward-backward method [11] and the gradient method [12]. However, these methods start with an initial guess and finally converge to a local optimum in practice. Few methods can escape from the local optimum to obtain the global optimum.

The tabu search algorithm [13] is a generalized heuristic global search technique with short-term or long-term memory, and it is suitable for solving many nonlinear opti-

Received March 12, 2001; revised April 25, 2003; accepted July 14, 2003.
Communicated by C. C. Jay Kuo.

mization problems. Tabu search approaches have been applied to codebook design of vector quantization [14], channel noise reduction [15], cell planning for mobile communications [16], alarm processing in power systems [17], and condensed nearest neighbor clustering [18]. The basic idea of the tabu search approach is to use adaptive memory structures and associated strategies to explore the search space of feasible solutions by means of a sequence of moves. The simplest version of this method only makes use of its short-term memory process. The elements of the move from the current solution to its selected neighbor are partially or completely recorded in the tabu list to forbid reversal of the replacement in future iterations. Without this assurance, the search would cycle between the first encountered local optimum and its neighbor. Although longer-term memory, which makes use of frequency information and more advanced principles, is generally essential to implement the most effective forms of tabu search, in some instances, considerable success is achieved using only short-term memory.

In this paper, a short-term memory version of the tabu search algorithm is applied to HMM training to search out the optimal structure of HMM for automatic speech recognition. The proposed TS-HMM training provides a mechanism that allows the search process to escape from a local optimum and to obtain a near global optimum.

In section 2 of this paper, the definition of HMM is given, and our proposed tabu search algorithm is described in section 3. The TS-HMM training algorithm is presented in section 4. Simulation results are given in section 5, and conclusions are drawn in section 6.

2. HIDDEN MARKOV MODEL

HMM is a probability model used to represent the statistic property of the stochastic process and is characterized by model parameters. The stochastic process in speech recognition consists of finite-length stochastic sequences called observation symbols, denoted by $\mathbf{O} = \mathbf{o}_1 \mathbf{o}_2 \dots \mathbf{o}_M$, where M is the dimension of the observation symbol. One HMM with N states (S_1, S_2, \dots, S_N) can be characterized by the parameter set $\lambda = \{\boldsymbol{\pi}, \mathbf{A}, \mathbf{B}\}$, where

- (1) $\boldsymbol{\pi} = [\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \dots, \boldsymbol{\pi}_N]$ is the initial distribution. It is used to describe the probability distribution of the observation symbol in the initial moment when $t = 1$, that is,

$$\boldsymbol{\pi} = P(q_1 = S_i) \quad i = 1, 2, \dots, N, \quad (1)$$

$$\text{which must satisfy } \sum_{i=1}^N \boldsymbol{\pi}_i = 1. \quad (2)$$

- (2) $\mathbf{A} = \{a_{ij} \mid i, j = 1, 2, \dots, N\}$ is the transition probability distribution matrix. Its element at row i , column j is the probability a_{ij} of transition from the current state i to the next state j , that is,

$$a_{ij} = P(q_t = S_j \mid q_{t-1} = S_i), \quad (3)$$

which must satisfy the following condition:

$$\sum_{i=1}^M \alpha_{ij} = 1. \quad (4)$$

- (3) $\mathbf{B} = \{b_{ik} \mid i = 1, 2, \dots, N, k = 1, 2, \dots, M\}$ is the observation symbol probability distribution matrix in the discrete HMM. Its element at row i , column k is the probability a_{ik} of the observation symbol with index k emitted by the current state i , and it must satisfy the following condition:

$$\sum_{k=1}^M b_{ik} = 1. \quad (5)$$

As mentioned above, HMM is used to approximate the probability of each observation symbol existing in the current state. When $\boldsymbol{\pi}$, \mathbf{A} , and \mathbf{B} are given, the probability $P(\mathbf{O} \mid \boldsymbol{\lambda})$ of the HMM system generating one random observation symbol can be calculated. Three essential problems of HMM must be solved:

- (i) how to effectively calculate the probability $P(\mathbf{O} \mid \boldsymbol{\lambda})$;
- (ii) how to select the optimal state sequence when the model $\boldsymbol{\lambda}$ is given;
- (iii) how to adjust the model parameter $\boldsymbol{\lambda}$ to make the probability $P(\mathbf{O} \mid \boldsymbol{\lambda})$ higher.

The Forward algorithm is often employed to solve the first problem, and the Viterbi algorithm is often used to solve the second one. To solve the third problem, the Gradient algorithm is employed. This paper focuses on solving the last problem. We use the tabu search algorithm to search the optimal model parameters $\boldsymbol{\lambda}$.

3. THE TABU SEARCH ALGORITHM

The tabu search algorithm, which was proposed by Glover and has been described in detail by Glover and Laguna [12], is a generalized heuristic global search technique. The simplified variant we investigate here utilizes short-time memory to forbid certain search directions at the present iteration of a sequential move selection in order to avoid cycling and escape from local optima. The elements of a move from the current solution to its selected neighbor are partially or completely recorded in the tabu list for the purpose of forbidding reversal of the replacement in a number of future iterations.

Starting from an initial solution (or a collection of such solutions), the tabu search approach examines neighbor solutions generated strategically or randomly and calculates them by computing their corresponding objective function values. If the best of these neighbor solutions is not tabu or if it is tabu but satisfies an aspiration criterion, then this solution is selected to be the new current solution to generate neighbor solutions for the next iteration. A simple aspiration criterion allows a tabu solution to be chosen regardless of its tabu status if its objective value is better than the best value of all the solutions so far found. The short-term instance of the tabu search algorithm is given as follows:

Tabu Search Algorithm ()

```

{
  generate initial solutions;
  calculate the current solution and the best solution;
  while the termination criterion has not yet been reached
  {
    generate the neighboring solutions for the solution currently being examined;
    calculate the corresponding objective values;
    select a neighbor to become the current solution and update the best solution;
    update the tabu list;
  }
}

```

4. THE TS-HMM ALGORITHM

In this paper, the configuration of HMM is a five-state left-right model, and the speech feature vectors are vector quantized into a codebook with a size of 256. Therefore, \mathbf{A} is a 5-by-5 matrix, and \mathbf{B} is a 5-by-256 matrix. As shown in Fig. 1, this model can represent speech signals whose properties change over time in a successive manner.

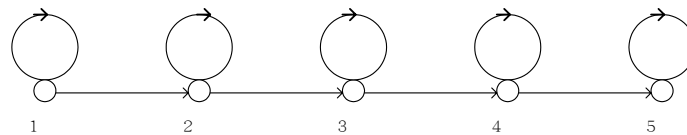


Fig. 1. A five-state left-right model.

Due to the configuration of the model, some transitions between states do not exist, so the corresponding elements in matrix \mathbf{A} are constantly zero, and these elements will not be encoded when search is performed. The optimal model parameters for the search problem must be mapped to the tabu search algorithm before it can be used. The mapping procedure is described below.

In TS-HMM training, the model is encoded into a string of real numbers between 0 and 1, which of course satisfy Eqs. (4) and (5). As shown in Fig. 2, this string is composed of two parts: \mathbf{SA} and \mathbf{SB} . These two parts are composed of the rows of matrices \mathbf{A} and \mathbf{B} , respectively. A solution of this algorithm is defined as s_l , consisting of a set of real numbers like the one shown in Fig. 2. The probability $p_n(\mathbf{O} | \lambda_n)$ of the HMM solution λ_n that generates the training observation sequences $\mathbf{O} = o_1 o_2 \dots o_M$ must be calculated as the objective function value.

The initial test solutions are generated randomly. After the first iteration, the test solutions are generated from the best solution of the current iteration by swapping two indices randomly. The tabu list memory stores the swapped indices only. It is a tabu condition if the swapped indices used to generate the new test solution from the best solution of the current iteration are the same as any records in the tabu list memory.

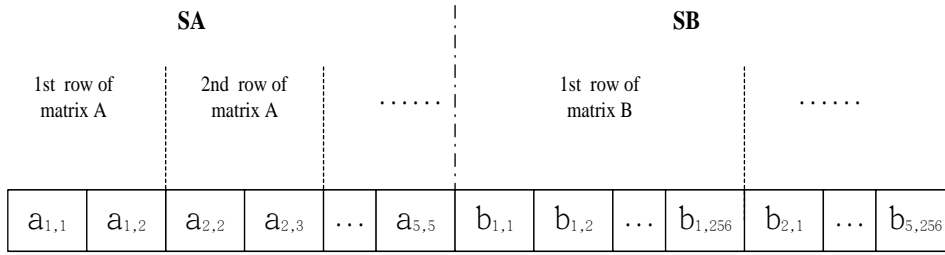


Fig. 2. The string representation of HMM.

Let $\theta = \{\lambda_1, \lambda_2, \dots, \lambda_{N_s}\}$ be the set of the test solutions; let $\lambda_c = \{\lambda_c(1), \lambda_c(2), \dots, \lambda_c(N)\}$ and $\lambda_b = \{\lambda_b(1), \lambda_b(2), \dots, \lambda_b(N)\}$ be the best solution of current iteration and the best solution of all iterations, respectively; let $V_i = \{v_1, v_2, \dots, v_{N_s}\}$, v_c , and v_b denote the set of objective function values for the test solutions, the objective function value for the best solution of the current iteration, and the objective function value for the best solution of all iterations, respectively, where v_l is the objective function value for solution λ_l , $1 \leq l \leq N_s$. The algorithm is given as follows:

- Step 0:** Set the tabu list size T_s , the number of test solutions N_s , and the optimum number of iterations l_m . Set the iteration counter $i = 1$ and insertion point of the tabu list $t_l = 1$. Generate N_s solutions $\theta = \{\lambda_1, \lambda_2, \dots, \lambda_{N_s}\}$ randomly, calculate the corresponding objective values $V_i = \{v_1, v_2, \dots, v_{N_s}\}$, and find the current best solution $\lambda_c = \lambda_j$, $j = \arg \max(v_l)$, $1 \leq l \leq N_s$. Set $\lambda_b = \lambda_c$ and $v_b = v_c$.
- Step 1:** Copy the current best path λ_c to each test solution λ_l , $1 \leq l \leq N_s$. For each test solution λ_l , $1 \leq l \leq N_s$, generate two random integers r_1 and r_2 , $1 \leq r_1 \leq N$, $1 \leq r_2 \leq N$, $r_1 \neq r_2$. Generate new test solutions as neighbors of the current solution by swapping $\lambda_l(r_1)$ and $\lambda_l(r_2)$. Calculate the corresponding objective values v_1, v_2, \dots, v_{N_s} for the new neighbor solutions.
- Step 2:** Sort v_1, v_2, \dots, v_{N_s} in increasing order. From the best neighbor solution to the worst neighbor solution, if the solution is a non-tabu solution or it is a tabu solution but its objective value is less than the best value of all iterations v_b (aspiration level), then choose this solution as the current best solution λ_c , choose its objective value as the current best objective value v_c , and go to step 3; otherwise, try the next neighbor solution. If all neighbor solutions are tabu solutions, then go to step 1.
- Step 3:** If $v_c < v_b$, set $\lambda_c = \lambda_b$ and $v_c = v_b$. Insert the swapped indices of the current best solution λ_c into the tabu list. Set the insertion point of the tabu list $t_l = t_l + 1$. If $t_l > T_s$, set $t_l = 1$. If $i < l_m$, set $i = i + 1$ and go to step 1; otherwise, record the best path index and terminate the algorithm.

5. SIMULATIONS

Ten experiments were conducted to validate the algorithm proposed in this paper. We recorded each word's pronunciation 10 times. Then we obtained 100 training obser-

vation sequences. For each word, we used the simple short-term memory of the tabu search algorithm and the forward-backward algorithm to train the HMM, and then obtained two sets of HMM model parameters and compared them. In this study, the length of the tabu list T_s was set to 20, the threshold of probability P_{th} was 0.17%, the number of iterations I_m was 800, and the number of solutions in each iteration N_s was 20. The initial model parameters were created randomly and were normalized to satisfy Eqs. (4) and (5).

In each experiment, the HMM training using the forward-backward algorithm was terminated when the average log probability increased less than 0.00001, and the TS-HMM training was terminated after 800 iterations.

We compared the HMMs trained by the tabu search algorithm and the forward-backward algorithm, respectively. Simulation results are shown in Table 1. They are divided into two parts: P_s and P_d . P_s denotes the average log probability of the HMM generated by the 10 training observation sequences of this HMM, and P_d denotes the average log probability of the HMM generated by the other 90 training observation sequences of the other HMMs.

Table 1. A comparison of the average log probability values obtained using the two algorithms.

Experiment	TS		Forward-Backward	
	P_s	P_d	P_s	P_d
#1	-3.3463	-9.0765	-4.2946	-8.9249
#2	-4.8036	-9.4363	-4.8116	-8.4035
#3	-4.6056	-8.4663	-5.6599	-8.3756
#4	-3.5379	-8.3139	-4.3562	-7.9967
#5	-4.6579	-9.9391	-5.1033	-7.6877
#6	-4.5324	-9.3661	-4.3394	-8.6031
#7	-3.2752	-9.3218	-4.7167	-8.4162
#8	-3.6225	-9.3123	-4.3607	-8.2275
#9	-3.8032	-9.6469	-4.5107	-9.2521
#10	-4.3190	-8.2152	-4.4864	-7.8755

As shown in Table 1, the HMMs trained by our simple instance of the tabu search algorithm had higher average log probabilities than the HMMs trained by the forward-backward algorithm except in experiment #6. Thus, the HMMs trained by the simple tabu search algorithm could better describe and recognize the training observation sequences.

6. CONCLUSIONS

This paper has proposed a TS-HMM training method. A simple version of the tabu search algorithm has been employed to repair the HMM model parameters λ and make

$P_n(\mathbf{O} | \lambda)$ highest. The simulation results indicate that TS-HMM training performed using this algorithm has a higher probability of finding globally optimal parameters and, in general, achieves better performance than it does when the forward-backward algorithm is employed. We observe that parallel implementation of the TS algorithm can be employed to reduce the search time.

REFERENCES

1. L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," in *Proceeding of IEEE*, Vol. 77, 1989, pp. 257-285.
2. H. Y. Gu, C. Y. Tseng, and L. S. Lee, "Isolated-utterance speech recognition using hidden Markov models with bounded state durations," *IEEE Transactions on Signal Processing*, Vol. 39, 1991, pp. 1743-1752.
3. K. Y. Su and C. H. Lee, "Speech recognition using weighted HMM and subspace projection approaches," *IEEE Transactions on Speech and Audio Processing*, Vol. 2, 1994, pp. 69-79.
4. J. T. Chien and H. C. Wang, "Adaptation of hidden Markov model for telephone speech recognition and speaker adaptation," in *IEE Proceedings of Vision Image Processing*, Vol. 144, 1997, pp. 129-169.
5. J. T. Chien, C. H. Lee, and H. C. Wang, "A hybrid algorithm for speaker adaptation using MAP transformation and adaptation," *IEEE Signal Processing Letters*, Vol. 4, 1997, pp. 167-135.
6. S. Wang and Y. Zhao, "Online Bayesian tree-structured transformation of HMMs with optimal model selection for speaker adaptation," *IEEE Transactions on Speech and Audio Processing*, Vol. 9, 2001, pp. 663-677.
7. L. M. Lee and H. C. Wang, "Representation of hidden Markov model for noise adaptive speech recognition," *Electronics Letters*, Vol. 31, 1995, pp. 616-617.
8. T. H. Hwang, L. M. Lee, and H. C. Wang, "Cepstral behaviour due to additive noise and a compensation scheme for noisy speech recognition," in *IEE Proceedings Vision Image Processing*, Vol. 145, 1998, pp. 316-321.
9. J. Picone, "Continuous speech recognition using hidden Markov models," *IEEE ASSP Magazine*, Vol. 7, 1990, pp. 26-41.
10. D. Burshtein, "Robust parametric modeling of duration in hidden Markov models," *IEEE Transactions on Speech and Audio Processing*, Vol. 4, 1996, pp. 240-242.
11. F. Jelinek, "Continuous speech recognition by statistical methods," in *Proceedings of IEEE*, Vol. 64, 1976, pp. 532-556.
12. 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," *The Bell System Technical Journal*, 1983, pp. 1035-1074.
13. F. Glover and M. Laguna, *Tabu Search*, Kluwer Academic Publishers, 1997.
14. H. C. Huang, S. C. Chu, J. S. Pan, and Z. M. Lu, "A tabu search based maximum descent algorithm for VQ codebook design," *Journal of Information Science and Engineering*, Vol. 17, 2001, pp. 753-762.
15. J. S. Pan and S. C. Chu, "Non-redundant VQ channel coding using tabu search strategy," *Electronics Letters*, Vol. 32, 1996, pp. 1545-1546.

16. C. Y. Lee and H. G. Kang, "Cell planning with capacity expansion in mobile communications: a tabu search approach," *IEEE Transactions on Vehicular Technology*, Vol. 49, pp. 1678-1691.
17. E. S. Wen and C. S. Chang, "Tabu search approach to alarm processing in power systems," in *IEE Proceedings of Generation, Transmission and Distribution*, Vol. 17, 1997, pp. 753-762.
18. V. Cerveron and F. J. Ferri, "Another move toward the minimum consistent subset: a tabu search approach to the condensed nearest neighbor rule," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, Vol. 31, 2001, pp. 408-413.



Tsong-Yi Chen (陳聰毅) received the B.S. degree from Tamkang University, and the M.S. degree and the Ph.D. degree from the Illinois Institute of Technology, Chicago, U.S.A. He is an Assistant Professor in the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences. His technical interests include data mining and knowledge discovery, schema versioning, and expert systems.



Xiao-Dan Mei (梅曉丹) received the Ph.D. degree from the Harbin Institute of Technology, Harbin, China in 2002. Her current research interests include pattern recognition and speech processing.



Jeng-Shyang Pan (潘正祥) received the B.S. degree in Electronic Engineering from the National Taiwan Institute of Technology, Taiwan, in 1986, the M.S. degree in Communication Engineering from National Chiao Tung University, Taiwan, in 1988, and the Ph.D. degree in Electronic Engineering from the University of Edinburgh, U.K., in 1996. Currently, he is a Professor in the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Taiwan. His current research interests include pattern recognition, speech coding and image processing.



Sheng-He Sun (孫聖和) was born in 1937. He is a professor at the Harbin Institute of Technology and a fellow of the Chinese Institute of Electronics. He has published more than 150 papers. His current research interests include electronic measurement and instrumentation, signal processing and system identification, data compression and pattern recognition.