A Heuristic Optimization Approach for HMM in Machinery Prognostics and Health Management

Wen-Zhu Liao †
Department of Industrial Engineering
Chongqing University, Chongqing, China
Tel: (+86) 13629771031, Email: liaowz@cqu.edu.cn

Shi-Hao Cui
Department of Industrial Engineering
Chongqing University, Chongqing, China
Tel: (+86) 13629771031, Email: 20150702070@cqu.edu.cn

Dan Li
Department of Industrial Engineering
Chongqing University, Chongqing, China
Tel: (+86) 13629771031, Email: ie_lidan@cqu.edu.cn

Abstract. This paper proposes a heuristic optimization approach for hidden Markov model (HMM) in machinery diagnosis based on simulated annealing (SA) algorithm and expectation maximization (EM) algorithm. As traditional HMM is sensitive to initial values and EM is easy to trap into a local optimization, SA is combined to improve HMM which can overcome local optimization searching problem. The proposed HMM has strong ability of global convergence, and optimizes the process of parameters estimation. Finally, through a case study, the computation results illustrate this SAEM-HMM has high efficiency and accuracy, which could help machinery diagnosis in practice.

Keywords: Hidden Markov model, Expectation maximization, Simulated annealing, Diagnosis

1. INTRODUCTION

Nowadays, HMM has been widely used in machinery prognostics and health management, due to its good performance about model construction and pattern classification for dynamic time sequence (Deng et al., 2013). However, previous HMM-based machinery diagnosis research focuses on applications of basic HMM theory mainly including two parts: improving HMM to better fit machine degradation in real production process, such as semi-HMM (Dong and He, 2007), multi branch-HMM (Thanh et al., 2014), etc; using feature extraction method for traditional HMM to achieve accurate diagnosis results (Xu et al., 2015; Zhou et al., 2015). However, these research cannot avoid that the traditional HMM is sensitive to initial value. Moreover, the traditional training algorithm is easy to trap into a local optimization.

Nowadays, some research has begun to consider the optimization of HMM training algorithm, which is focused on human behavior recognition, speech recognition (Huda et al., 2014), visual speech recognition (Lee and Park, 2006), etc. Some heuristic algorithms are studied to search global optimum solution such as genetic algorithm (GA) (Liang, 2013) and Tabu search (TS), but general application reflects their performance to a certain extent. Thus, HMM training algorithm needs to be further studied.

SA which origins from a physical annealing process can solve the global optimization problem and get better solution. This paper is devoted to optimize HMM process by combining SA with EM. The local optimization problem of model parameters estimated by EM can be overcome by SA. Moreover, it can avoid the sensitivity of initial values selection.

2. HIDDEN MARKOV MODEL

HMM presents a double stochastic process based on general Markov chain. In general, it can be formulated as $\lambda = (N,M,\boldsymbol{\pi},\boldsymbol{A},\boldsymbol{B})$. Its parameters are defined as follow.

1) N is the states number of hidden Markov chain, where the states are $\theta_1,\theta_2,\cdots,\theta_N$. Suppose the state of hidden Markov chain at time t is recorded as q_t .
2) M is the possible number of observation values for each state, where the observation values are $v_1,v_2,\cdots,v_M$. Suppose the observation value at time t is recorded as o_t .
3) $\pi = (\pi_1,\pi_2,\cdots,\pi_N)$ is the probability distribution of initial states, where

$$\begin{cases} \pi_i = P(q_1 = \theta_i) \\ \sum_{i=1}^{N} \pi_i = 1 \end{cases} \quad (1)$$
$$1 \le i \le N$$

4) $\boldsymbol{A}=(a_{ij})_{N\times N}$ is the probability matrix of states transition, and a_{ij} represents the transition probability from state i to state j, shown as

$$\begin{cases} a_{ij} = P(q_{t+1} = \theta_j \mid q_t = \theta_i) \\ \sum_{j=1}^{N} a_{ij} = 1 \end{cases} \quad (2)$$
$$1 \le i$$
$$j \le N$$

5) $\boldsymbol{B}=(b_{jk})_{N\times M}$ is the probability matrix of observation values, and b_{jk} represents the probability of observation value v_k at time t for state j, shown as

$$\begin{cases} b_{jk} = P(o_t = v_k \mid q_t = \theta_j) \\ \sum_{k=1}^{M} b_{jk} = 1 \end{cases} \quad (3)$$
$$1 \le j \le N$$
$$1 \le k \le M$$

The above HMM can be abbreviated as $\lambda = (\boldsymbol{\pi},\boldsymbol{A},\boldsymbol{B})$ and EM algorithm is conventionally used to train HMM. The EM algorithm is used to search the maximum likelihood value of HMM parameters. The parameters of HMM can be reevaluated, shown as

$$\overline{\pi}_i = \gamma_1(i) \quad (4)$$

$$\overline{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (5)$$

$$\overline{b}_{jk} = \frac{\sum_{t=1,o_t=v_k}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)} \quad (6)$$

where $\overline{\pi}_i$, $\overline{a}_{ij}$, $\overline{b}_{jk}$ are the model parameters of $\overline{\lambda}$. The forward probability variable $\alpha_t(i)$, the backward probability variable $\beta_t(i)$ and the probability variable $\gamma_t(i)$ are defined as follows:

$$\alpha_t(i) = p(O_1,O_2,\cdots,O_t,q_t = \theta_i \mid \lambda) \quad (7)$$

$$\beta_t(i) = p(O_{t+1},O_{t+2},\cdots,O_T \mid q_t = \theta_i,\lambda) \quad (8)$$

$$\gamma_t(i) = p(q_t = \theta_i \mid O,\lambda) \quad (9)$$

where $t = 1,2,\cdots,T$ and $i = 1,2,\cdots,N$.

And $\xi(i,j)$ is the probability of machine state in θ_j at time $t+1$ while machine state is θ_i at time t , shown as

$$\xi(i,j) = p(q_t = \theta_i, q_{t+1} = \theta_j \mid O,\lambda) \quad (10)$$
$$1 \le t \le T-1$$

Hence, there are,

$$\alpha_t(i) = \begin{cases} \pi_i b_{i1} & , \; t = 1, 1 \le i \le N \\ \left\{ \sum_{j=1}^{N} a_{ij}\alpha_{t-1}(j) \right\} b_{j(t+1)} & , \; 2 \le t \le t-1, 1 \le i \le N \end{cases} \quad (11)$$

$$\beta_t(i) = \begin{cases} 1 & , \; t = T, 1 \le i \le N \\ \sum_{j=1}^{N} a_{ij} b_{j(t+1)} \beta_{t+1}(j) & , \; 1 \le t \le T-1, 1 \le i \le N \end{cases} \quad (12)$$

$$P(O \mid \lambda) = \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_t(i) a_{ij} b_{j(t+1)} \beta_{t+1}(j) \quad (13)$$

$$\xi(i,j) = \frac{\alpha_t(i) a_{ij} b_{j(t+1)} \beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_t(i) a_{ij} b_{j(t+1)} \beta_{t+1}(j)} \quad (14)$$

The new model is reconstructed to be $\overline{\lambda} = (\overline{\boldsymbol{\pi}},\overline{\boldsymbol{A}},\overline{\boldsymbol{B}})$. Through new iterations until $p(O \mid \overline{\lambda}) - p(O \mid \lambda) < \delta$, the optimal $\overline{\lambda}_{EM}$ is obtained, where δ (which generally is set to 10^{-4}) is predetermined convergent condition.

3. SAEM TRAINING ALGORITHM

As HMM trained by EM may trap into local optimization due to its initial values and SA can help to solve global optimization problem by using Metropolis criterion and controlling the deterioration process of temperature (Huda et al., 2009), a novel method combing SA and EM is developed in this paper. Metropolis criterion means when the parameter λ_{old} changes into λ_{new}, the system energy should be changed from $E(\lambda_{old})$ to $E(\lambda_{new})$ respectively. The acceptance probability of p is

$$p = \begin{cases} 1 & , \quad if\ E(\lambda_{new}) < E(\lambda_{old}) \\ \exp(-\dfrac{E(\lambda_{new}) - E(\lambda_{old})}{T_{(t)}}) , & if\ E(\lambda_{new}) > E(\lambda_{old}) \end{cases} \tag{15}$$

$$T_{(t)} = \frac{T_0}{\log 1 + t} \tag{16}$$

where T_0 is initial temperature given by empirical formula.

And $E(\lambda) = -\log P(O|\lambda)$ is selected to be energy function, thus the optimal parameter λ_{final} of HMM can be obtained by minimizing $E(\lambda)$. The training steps of this proposed heuristic algorithm are given as follows:

(i) Set initial values of the parameter λ for HMM and assume initial temperature T_0 to be the standard deviation of selected samples. Assume the times of temperature annealing is s_num=1 and the times of iteration is iter_num=1.

(ii) Take the parameter λ as λ_{old}. According to Eqs.(11)-(13), calculate $P(O|\lambda_{old})$. Then, $E(\lambda_{old})$ can be obtained by reversing its logarithm.

(iii) Take the reevaluated parameter $\bar{\lambda}$ obtained by Eqs.(4)-(6) as λ_{new}. According to Eqs.(11)-(13), calculate $P(O|\lambda_{new})$. Then, $E(\lambda_{new})$ can be obtained by reversing its logarithm.

(iv) If $E(\lambda_{new}) < E(\lambda_{old})$, the new solution is accepted and the times of temperature annealing is s_num=s_num+1. Otherwise, go to the next step.

(v) Calculate p by Eq.(15). And use Rand() function to generate a constant c less than 1 randomly. If $p > c$, the new solution is accepted. Otherwise, reject the new solution and the algorithm ends.

(vi) Address the annealing method by Eq.(16). If the terminal condition is satisfied, go to the next step. Otherwise, assume the times of iteration is iter_num=iter_num+1, then go to step (iii).

(vii) Return the final model $\bar{\lambda}_{final}$, End the algorithm.

4. MACHINERY DIAGNOSIS BASED ON PROPOSED HMM

Machinery diagnosis mainly includes three processes: data analysis, HMM construction and state evaluation. The observation sequence for HMM construction training could be processed by monitoring data of the machine. Based on optimal HMM parameters determined, the current state of the machine could be estimated by inputting real-time observation data.

As machine deterioration process is hard to describe, this paper classifies its health state into five levels: healthy, good, normal, unhealthy and broken, and then maps the relationship between health state and machine performance data so as to train and evaluate HMM. This paper optimizes HMM by using the SAEM algorithm. The optimal $\bar{\lambda}_{final}$ of the model could be obtained by the proposed optimization approach. Then the observation sequence given by the real-time collected data of the machine can be inputted into this evaluation model to estimate machine health state. The probability that the observation sequence is in state θ_i at time t can be calculated as

$$\bar{\gamma}_t(i) = \sum_{j=1}^{N} \bar{\xi}_t(i,j) = \frac{\bar{\beta}_t(i)\bar{\alpha}_t(i)}{P(O|\bar{\lambda}_{final})} \tag{17}$$

According to maximum membership principle, machine health state at time t calculated by the Eq.(17), which can be described by a column map. Therefore, one state curve used to describe machine health state at different time t could well depict machine degradation.

5. A CASE STUDY

This paper takes a drilling machine as a case study to apply SAEM-HMM in machinery diagnosis, in which its spindle voltage signal is used to indicate machine performance. According to the method of reference (Liao, 2011), 500 groups of spindle voltage signal are collected. Based on the data, machine health state is divided into 5 levels (as Table 1). After processing, machine health index can be obtained.

Table 1 Classification of machine health

Health state	Healthy	Good	Normal	Unhealthy	Broken
Health index	0.85~1	0.7~0.85	0.5~0.7	0.3~0.5	0~0.3

As this proposed approach focuses on the sensitivity to initial values of traditional Markov model, therefore model parameters can be randomly given. Then machine health index considered as the observation sequence is taken into EM algorithm and SAEM algorithm for training, $\bar{\lambda}_{EM} = (\bar{\boldsymbol{\pi}}_{EM}, \bar{\boldsymbol{A}}_{EM}, \bar{\boldsymbol{B}}_{EM})$ and $\bar{\lambda}_{final} = (\bar{\boldsymbol{\pi}}_{final}, \bar{\boldsymbol{A}}_{final}, \bar{\boldsymbol{B}}_{final})$ can be obtained respectively. In order to validate this proposed SAEM-HMM, 50 groups of data (seen in Table 2) are selected to calculate the probability while the machine in different states.

Then, machine health state could be estimated by maximum membership principle. The comparison of results is shown as Table 3, which can prove this proposed algorithm achieve much higher accuracy.

Table 2: The evaluation results of machine health states.

	EM algorithm					SAEM algorithm				
Index \ State	Healthy	Good	Normal	Unhealthy	Broken	Healthy	Good	Normal	Unhealthy	Broken
0.89	0.79	0.18	0.03	0	0	0.82	0.18	0	0	0
0.87	0.57	0.42	0.01	0	0	0.63	0.37	0	0	0
0.85	0.51	0.47	0.01	0	0	0.55	0.446	0.004	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.62	0.16	0.36	0.48	0.06	0	0.05	0.465	0.464	0.021	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.41	0	0	0.02	0.68	0.30	0	0.004	0.026	0.65	0.32
0.30	0	0	0.01	0.53	0.46	0	0.007	0.013	0.37	0.61

Table 3 Comparison of results

	EM algorithm	SAEM algorithm
Health index \ Evaluation results	Right=1 & Wrong=0	Right=1 & Wrong=0
0.89	1	1
⋮	⋮	⋮
0.74	0	1
⋮	⋮	⋮
0.62	1	0
⋮	⋮	⋮
0.53	0	1
⋮	⋮	⋮
0.30	0	1
Rate of accuracy	47/50	49/50

6. CONCLUSION

In this paper, a heuristic algorithm considering SA and EA is developed to optimize HMM for machinery diagnosis. With the advantage of global convergence by SA, the local optimization searching problem of model parameters estimated by EM can be overcome. Moreover, it avoids the issue of initial values selection. Through a case study, the computation results indicate that this proposed algorithm can achieve high accuracy, which makes this new approach efficient and applicable in machinery diagnosis. Later, in order to achieve good machinery prognostics and health management, the future work will focus on the extension of this optimization model.

ACKNOWLEDGMENTS

The authors would like to thank anonymous referees for their remarkable comments and great support by National Natural Science Foundation of China (71301176) and Research Fund for the Doctoral Program of Higher Education of China (No. 20130191120001)

REFERENCES

Deng, C., Sun, Y., Li, R., Wang, Y., and Xiong, Y. (2013) Hidden markov model based on the heavy-duty CNC health state estimate. *Computer Integrated Manufacturing Systems*, *19*(3), 552-558.

Dong, M., and He, D. (2007) A segmental hidden semi-Markov model (HSMM) based diagnostics and prognostics framework and methodology. *Mechanical Systems and Signal Processing*, *21*(5), 2248-2266.

Huda, S., Yearwood, J.,and Togneri, R. (2009) A stochastic version of Expectation Maximization algorithm for better estimation of Hidden Markov Model. *Pattern Recognition Letters*, 30, 1301-1309.

Huda, S., Yearwood, J., and Togneri, R. (2014) Hybrid metaheuristic approaches to the expectation maximization for estimation of the hidden markov model for signal modeling. *IEEE Transactions on Cybernetics*, 44(10), 1962-1977.

Lee, J. and Park, C. (2006) Training Hidden Markov Models by Hybrid Simulated Annealing for Visual Speech Recognition. Systems, Man and Cybernetics, 2006. *SMC '06. IEEE International Conference on*, 1, 198-202.

Liang, P. (2013) Human behavior recognition based on GA-HMM. *Electronic test (Z2)*, 88-90.

Liao, W. (2011) Research of predictive maintenance policy and integrated production scheduling model based on machine degradation [D]. Shanghai: *Shanghai Jiao Tong University*.

Thanh, T., Florent, C., and Christophe, B. (2014) Hidden Markov Models for diagnostics and prognostics of systems under multiple deterioration modes. *European Safety and Reliability Conference-ESREL 2014*, 1197-1204.

Xu, H., Fan, Y., Wu, J., Gao, Y., and Yu, Z. (2015) Bearing fault diagnosis method based on singular value decomposition and hidden Markov model. *Control and Decision Conference (CCDC)*, 2015 27th. IEEE, 6355-6359.

Zhou, H., Chen, J., Dong, G., and Wang, R. (2015) Detection and diagnosis of bearing faults using shift-invariant dictionary learning and hidden markov model. *Mechanical Systems & Signal Processing*, 72, 65-79.