CHEMICAL ENGINEERING TRANSACTIONS

VOL. 33, 2013

Guest Editors: Enrico Zio, Piero Baraldi

ISBN 978-88-95608-24-2; ISSN 1974-9791

A publication of

AIDIC

The Italian Association of Chemical Engineering
Online at: www.aidic.it/cet

DOI: 10.3303/CET1333059

Performance Data Prognostics Based on Relevance Vector Machine and Particle Filter

Yang Hu*[a], Pengcheng Luo[b]

[a] Energy Department, Politecnico di Milano, Via Ponzio 34/3, 20133 Milan, Italy
[b] Coll. of Information System and Management, National Univ. of Defense Technology, Changsha 410073, China
hy4981326@gmail.com

Prognostics of equipment state is one of the most important and difficult issue in Prognostics and Health Management (PHM) technology. In this paper, Relevance Vector Machine (RVM) and Particle Filter (PF) are used to analyse equipment state and predict its trend. RVM regression is used to extract the trend information of historical measurement data (trajectories). Then RVM based PF method is proposed to figure out the particle sampling distribution using RVM regression, also a new weight calculation method of particle is introduced based on the likelihood sum of whole trajectories. The proposed RVM based PF method can handle the situation that state-transition function of equipment state is unavailable, and it can provide satisfied prognostics of equipment state.

1. Introduction

In the Prognostics and Health Management (PHM), the prognostics of equipment state is the most important and difficult issue. If the information about the equipment state and its future evolution is available, it is possible for us to know when the equipment will lose function and opportunely schedule the maintenance plans at the most convenient and inexpensive times (Zio E., 2013). Motivated to achieve such a goal, an efficient and accurate computational algorithm is strongly needed to obtain the prognostics of equipment state using historical available data (Said A. Said, et al. 2011). Relevance Vector Machine (RVM) and Particle Filter (PF) are two highlighted prognostic methods which are accepted widely in the world.

RVM is first introduced by Michael (Tipping M E., 2001). This model combines the theories of Markov, Bayesian, Automatic Relevance Determination and Maximum Likelihood, and obtains a sparse sum of kernel function to solve the problems of regression and classification. Compared to the Support Vector Machine (SVM), RVM has a significant reduction in the cost of computational complexity and consumption (Liu Xuemei, et al. 2010). In addition, the kernel function in RVM needs not to satisfy the Mercer condition, also a noise factor is added into the model expression to describe the uncertainty. Many scholars use RVM in different prognostic applications and achieve good results. In recent years, some improvements of RVM have been made to enhance the robustness of traditional RVM and make it more accurate when proceeding the big- noise data (Han and Zhao, 2011).

PF based prognostic methods have been well studied in the applications of fault and remaining useful life prediction. It has a good adaption of non-linear and non-Gaussian data, but there are also some problems like particle degradation and impoverishment (Yang, et al. 2008), which decrease the accuracy of the prediction (Shenoy, et al. 2012). These two problems could be solved by choosing a good particle sampling distribution. Usually the state-transition function is used to figure out this distribution, but this is not always working well because of neglecting the information of measurement data. What's more, in some situations, the state-transition function is not available.

In some prognostic applications, only the measurement data of equipment state are available. In such cases, PF cannot work on it, but we can use RVM regression to generate the sampling distribution for PF. RVM regression can extract the evolution trend of equipment state from measurement data, and provide highly accurate interval estimation of equipment state. So this interval is a good choice of PF's sampling

Please cite this article as: Hu Y., Luo P., 2013, Performance data prognostics based on relevance vector machine and particle filer, Chemical Engineering Transactions, 33, 349-354 DOI: 10.3303/CET1333059

distribution. In this paper, this RVM based PF method is proposed to solve the prognostic problem in which state-transition function is not available. The remainder of this paper is organized as follows: section 2 introduces the basic theories of RVM and PF; section 3 proposes the RVM based PF method, which contains the sampling distribution and particle's weight calculation; test and verification of RVM based PF are made in section 4, also the comparison of two different weight calculation methods; section 5 concludes the whole paper.

2. Basic theories of RVM and PF

RVM and PF are two of the most useful tools to do the regression and prognostic work. They are all under the framework of Bayesian theory. RVM can provide a sparse weighted sum of kernel function as the regression of equipment state. PF makes use of system state-transition function and observation data to update the posterior distribution of equipment state. The output of RVM is the foundation of RVM based PF.

2.1 RVM

The model expression of RVM is almost the same like SVM, but the weight vector is calculated using Bayesian theory. Given the training data $\{\mathbf{x}_i, y_i\}_{i=1}^{M}$, where $\mathbf{x}_i$ is the input variable vector , y_i is the target value, M is the length of training data, the RVM regression expression is:

$$y(\boldsymbol{x};\boldsymbol{w}) = \sum_{n=1}^{N} w_n k(\boldsymbol{x}, \boldsymbol{x}_n) + w_0,\quad u_i = y(\boldsymbol{x}_i;\boldsymbol{w}) + \varepsilon_i \quad i = 1, 2, ..., M \tag{1}$$

where $\boldsymbol{w} = [w_0, w_1, \cdots, w_N]$ is the weight vector, N is the number of relevance vectors (RV), $k(\boldsymbol{x}, \boldsymbol{x}_n)$ is the value of kernel function, $\varepsilon_i \sim N(0, \sigma^2)$ is the independent and identically distributed noise, u_i is the regression value of y_i.

If $\boldsymbol{w}$ and ε_i are directly estimated, there will be the problem of over-fitting, also the sparseness of model will not achieved. Therefore, the hyper-parameter α_i is introduced. Suppose that $\boldsymbol{w}$ obeys Gaussian distribution with zero-mean and variance α_i^{-1}, i.e.,

$$p(w_i \mid \alpha_i) \sim N(w_i \mid 0, \alpha_i^{-1}) \tag{2}$$

$$p(\boldsymbol{w} \mid \boldsymbol{\alpha}) = \prod_{i=0}^{N} N\left(w_i \mid 0, \alpha_i^{-1}\right) \tag{3}$$

where $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, ..., \alpha_N]$, $\boldsymbol{\Phi}$ is the matrix of kernel function.

$\boldsymbol{\alpha}$ and σ are the parameters need to be estimated, the updating process is as follows:

$$\alpha_i^{new} = \frac{\gamma_i}{\mu_i^2} \tag{4}$$

$$(\sigma^2)^{new} = \frac{\|\boldsymbol{u} - \boldsymbol{\Phi\mu}\|^2}{N - \boldsymbol{\Sigma}_i \gamma_i} \tag{5}$$

where $\boldsymbol{\Sigma} = (\sigma^{-2}\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \boldsymbol{A})^{-1}$, $\boldsymbol{\mu} = \sigma^{-2}\boldsymbol{\Sigma}\boldsymbol{\Phi}^T\boldsymbol{u}$, $\gamma_i = 1 - \alpha_i \Sigma_{ii}$.

In the process of updating, most of α_i will get close to infinity. The remaining α_i are sparse, and the corresponding w_i are the RVs. The details of RVM could be found in paper (Tipping, 2001).

2.2 PF

The basic theory of PF is performed in many academic papers. In this paper, only the five important procedures are discussed:

1. State transition and observation function

$$x_t = f(x_{t-1}, v_t) \tag{6}$$

$$z_t = h(x_t, \omega_t) \tag{7}$$

where t is the time step, x_t is system state, z_t is the measurement of x_t, v_t and ω_t are the noise of process and measurement, respectively.

2. Particle Generation
   Particles are generated from the sampling distribution, in many cases, $p(x_t \mid x_{t-1})$ which can be obtained by Eq (6) is used as the sampling distribution, for its easier acquisition.
3. Weight Update
   When new measurement is available, weight can be updated:

$$w_t^i = w_{t-1}^i p(z_t \mid x_t^i) \tag{8}$$

   w_t^i is the weight of particle i at time t, $p(z_t \mid x_t^i)$ can be obtained by Eq (7).
4. Resampling
   If many particles' weights are too small, it means that these particles are useless. A resampling is needed to get more effective particles. The concept of Effective Sample Size (ESS) can be used to describe the degradation of particle:

$$ESS_t = \left(\sum_{i=1}^{n} \left(w_t^i \right)^2 \right)^{-1} \tag{9}$$

where n is the number of particles, ESS_t reaches its max value n when $w_i = 1/n,\ i = 1, 2, ..., n$. If ESS_t is too small, the resample need to be taken. A threshold of T_{ESS} should be set before PF start, usually $T_{ESS} = n/2$, i.e., if $ESS_t < T_{ESS}$, the resampling need to be taken. The detail resampling process could be found in paper (Arulampalam, et al. 2002).
5. Prognosis
The posterior distribution of x_t can be calculated as:

$$p(x_t \mid z_{0:t}) \approx \sum_{i=1}^{n} w_t^i \delta(x_t - x_t^i) \tag{10}$$

If n is large enough, Eq (10) can effectively approximate the true value of $p(x_t \mid z_{0:t})$.

3. RVM based PF

PF provides a satisfied prognostic result of equipment state, but it needs the state-transition function like Eq (6). In the actual engineering application, state-transition function is not always available; sometimes we can only get the trajectory of equipment state. In this case, PF method cannot work. To overcome this, RVM regression is used to take the place of state-transition function and generate the particle. The RVM based PF is different from traditional PF in two aspects: Sampling Distribution and Particle Weight.

3.1 Sampling Distribution

According to subsection 2.1, if the trajectories of equipment state are available, a regression (weighted sum of kernel function) could be provided by RVM. At each time step, the regression value of equipment state obeys Gaussian distribution:

$$u_t \sim N\left(\sum_{i=1}^{N} w_n k(\boldsymbol{x}_t, \boldsymbol{x}_n) + w_0, \sigma^2 \right) \tag{11}$$

Due to the high accuracy of RVM regression, u_t is very close to the posterior distribution. In addition, it is very easy to sample. So u_t is an good sampling distribution for PF. Let $\{p_t^i\}$ denote the particles, where

$i=1,2,...,n,\ t=1,2,...,T$, n is the number of particles, T is the time length of equipment state's trajectory. So the sampling method is:

$$p_t^i \sim N\left(\sum_{i=1}^{N} w_n k(\boldsymbol{x}_t, \boldsymbol{x}_n) + w_0, \sigma^2\right) \tag{12}$$

The state-transition function is not needed any more, because u_t can be obtained by RVM regression at each time step.

3.2 Particle Weight

In the traditional PF, the weight is updated at each time step according to the likelihood of each particle. If the measurement noise is Gaussian noise, the weight is calculated as follows:

$$L\left(z_t \mid p_t^i\right) = \frac{1}{\omega_t\sqrt{2\pi}} \exp\left[-\frac{(z_t - p_t^i)^2}{2\omega_t^2}\right], i=1,2,...,n,$$

$$w_t^i = \frac{L\left(z_t \mid p_t^i\right)}{\sum_{i=1}^{n} L\left(z_t \mid p_t^i\right)}, i=1,2,...,n \tag{13}$$

where $L\left(z_t \mid p_t^i\right)$ is the likelihood of particle p_t^i, ω_t is measurement noise and w_t^i is the weight of p_t^i. However, for the trajectory of equipment state, this weight calculation method is not the best one. We should use the likelihood of whole trajectory to calculate the weight rather than a single time step. So Eq (13) could be improved as:

$$L_{sum}\left(z_t \mid p_t^i\right) = \sum_{t=1}^{T}\left(\frac{1}{\omega_t\sqrt{2\pi}} \exp\left[-\frac{(z_t - p_t^i)^2}{2\omega_t^2}\right]\right), i=1,2,...,n$$

$$w_t^{i\prime} = \frac{L_{sum}\left(z_t \mid p_t^i\right)}{\sum_{i=1}^{n} L_{sum}\left(z_t \mid p_t^i\right)}, i=1,2,...,n \tag{14}$$

where $L_{sum}\left(z_t \mid p_t^i\right)$ is the sum likelihood of trajectory, $w_t^{i\prime}$ is the improved weight of p_t^i.

The resampling and prognosis methods of RVM based PF are the same like traditional PF.

4. Test and verification

In this section, a series of simulated trajectories of equipment state is used to test and verify the RVM based PF. A comparison between two different weight calculation methods is made to show the effectiveness of RVM based PF.

4.1 Data Generation Method

The following model is used to generate the trajectories of equipment state:

$$y_t = y_0 - \exp(at) - R_t + N(0,\sigma)$$

$$R_t = \begin{cases} 0, \text{ if } v > p \\ bt, \text{ if } v \le p \end{cases} \tag{15}$$

y_t is the state at time t, y_0 is the initial value, a , b and σ are the parameters which needed to be set according to the actual application, v is a random number generated from uniform distribution $U(0,1)$.

In this model, there are two parts: exponential part $\exp(at)$ and the random part R_t. $\exp(at)$ represents the normal degradation process and R_t represents the random damage with the probability p .

In this paper, 200 trajectories are generated using Eq (15). The parameters' value and the plot of trajectories are shown in Table 1 and Figure 1, respectively:

Table 1: Value of each parameter

Parameter	a	b	p	σ
Value	0.02	0.01	0.25	0.5

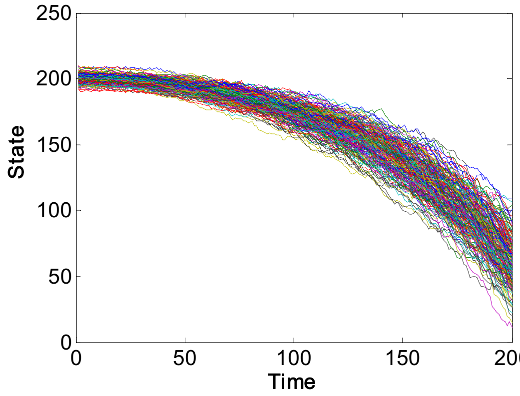

Figure 1: 200 trajectories of equipment state

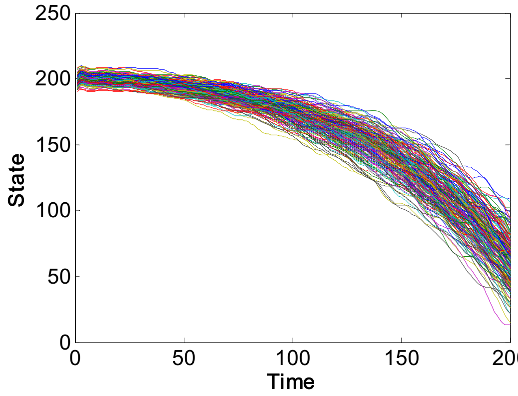

Figure 2: RVM regression of data

Figure 2 is the result of RVM regression of 200 equipment state trajectories. The trajectories in Figure 2 are smoother than the ones in Figure 1, mainly because RVM regression can restrain the measurement noise of data and is closer to the true value.

4.2 Prognostic Result

In this subsection, 200 equipment state trajectories are divided into two parts: the former 199 trajectories are used as the training data set, and the last one trajectory is used as the test set. Namely 199 particles (each trajectory is a particle) are available to predict the 200th trajectory. Figure 3 shows the predicted value and interval given the 1~50 time steps train data, while Figure 4 is the situation of 1~130 time steps training data.

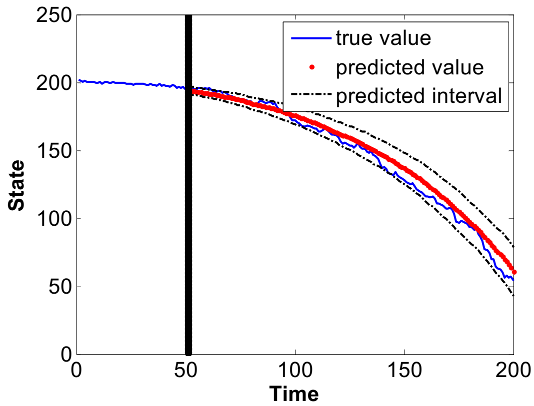

Figure 3: Prognostic result (training length is 50) using Eq (14)

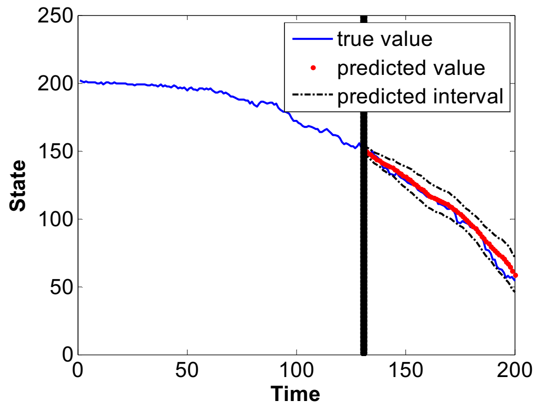

Figure 4: Prognostic result (training length is 130) using Eq (14)

To make a comparison of the different weight calculation methods in Eq (13) and Eq (14), we also use Eq (13) to figure out the prognostic results, which are shown in Figure 5 and 6. The Root-Mean-Square Error (RMSE) is used to evaluate the Prognostic results, shown in Table 2.

$RMSE(w)$ and $RMSE(w')$ represent the RMSEs using Eq (14) and Eq (13) respectively. We can see that the former ones are smaller than the later ones, especially when the training length is long. Because Eq (14) uses the sum of likelihood of training data to calculate the weights, when the training length is long, more information of likelihood is available. In this way, the likelihood is more accurate than the situation of short training length.

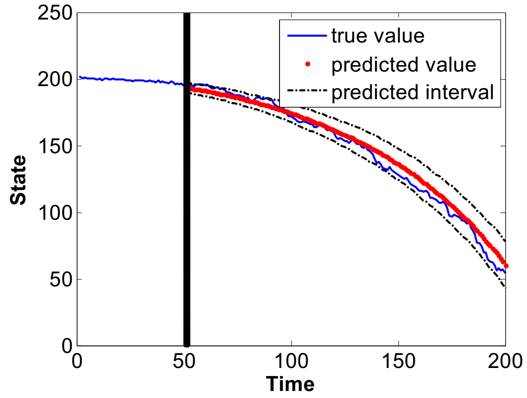

Figure 5: Prognostic result (training length is 50) using Eq *(13)*

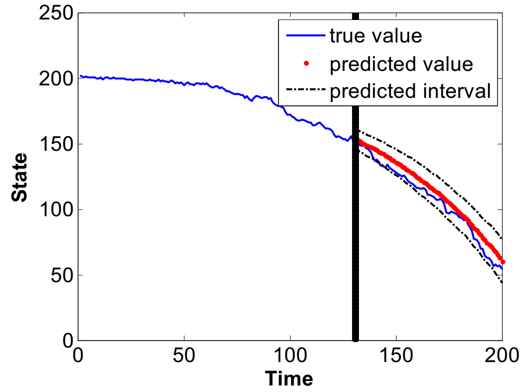

Figure 6: Prognostic result (training length is 130) using Eq *(13)*

Table 2: RMSE of two weight calculation methods

Train length	50	130
$RMSE(w)$	5.40	4.35
$RMSE(w')$	5.70	7.11

5. Conclusion

This paper proposes an RVM based PF prognostics method, which can handle the situation that state-transition function of equipment state is unavailable. RVM based PF uses RVM regression to figure out the sampling distribution, which makes the particles closer to the true value. A new weight calculation method of particle is presented, using the likelihood sum of whole trajectory rather than a single time step. Making full use of more trajectory likelihood information, this method is more accurate than the traditional one.

The prognostic method proposed in this paper needs a large number of measurement data (trajectories) to achieve good results (actually one trajectory means one particle). The number of data could be set according to the estimation accuracy. This is probably the main disadvantage of this method. How to efficiently and accurately generate the particles based on small number of measurement data (trajectories) is the future topic of this method.

References

Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002), A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking, IEEE transaction on signal processing, VOL. 50, NO. 2,

Han Min, Zhao Yao, 2011, Dynamic control model of BOF steelmaking process based on ANFIS and robust relevance vector machine, Expert Systems with Applications 38, 14786–14798

Said A. Said, Mansour Emtir, Iqbal M. Mujtaba, Neural Network Based Corrlations for Estimating the First and Second Dissociation Constant of Carbonic Acid in Seawater, Chemical Engineering Transactions 24, 2011

Shenoy A.V. ., Prakash J., Prasad V., Shah S.L., Auley K.B.Mc., 2012, Practical issues in state estimation using particle filters: Case studies with polymer reactors, Journal of Process Control,

Tipping M E., 2001, Sparse Bayesian Learning and the Relevance Vector Machine, Journal of Machine Learning Research 1, 211-244

Xuemei Liu, XiaoHui Zhang, Jin Yuan, 2010, Relevance vector machine and fuzzy system based multi-objective dynamic design optimization: A case study, Expert Systems with Applications 37, 3598–3604

Yang Xiaojun, Xing Keyi, Shi Kunlin, Pan Quan, 2008, Joint state and parameter estimation in particle filtering and stochastic optimization, Journal of Control Theory and Applications, 6 (2) 215–220

Zio E., 2013, Prognostics and Health Management of Industrial Equipment, Diagnostics and Prognostics of Engineering Systems: Methods and Techniques, Seifedine Kadry (Ed.) 333-356.