Novel Bayesian inference on optimal parameters of support vector machines and its application to industrial survey data classification

Jingjing Zhong, Peter W. Tse, Dong Wang*

Department of Systems Engineering and Engineering Management, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong, China

A R T I C L E   I N F O

Article history:
Received 1 September 2015
Received in revised form 28 December 2015
Accepted 28 December 2015
Available online 7 June 2016

Keywords:
Support vector machine
Particle filter
State space model
Cross-validation accuracy

A B S T R A C T

Engineering Asset Management (EAM) is a recently attractive discipline and it aims to address valuable contributions of asset management to organization's success. As of today, there is no specific method to evaluate performance of EAM standards. This paper aims to fill this gap and rank performance of asset management automatically after conducting survey, instead of evaluating questionnaires, analyzing results and ranking performances with a tedious process. Hence, it is necessary to develop intelligent data classification to simplify the whole procedure. Among many supervised learning methods, support vector machine attracts much attention for binary classification problems and its extension, namely multiple support vector machines, is able to solve multiclass classification problems. It is crucial to find optimal parameters of support vector machines prior to their use for prediction of unknown testing data sets. In this paper, novel Bayesian inference on optimal parameters of support vector machines is proposed. Firstly, a state space model is constructed to find the relationship between parameters of support vector machines and guess cross-validation accuracy. Here, the guess cross-validation accuracy aims to prevent support vector machines from overfitting. Secondly, particle filter is introduced to iteratively find posterior probability density functions of the parameters of support vector machines. Then, optimal parameters of support vector machines can be found from the posterior probability density functions. Ultimately, survey data collected from industry are used to validate the effectiveness of the proposed Bayesian inference method. Comparisons with some randomly selected parameters are conducted to highlight the superiority of the proposed method. The results show that the proposed Bayesian inference method can result in both high training and testing accuracies.

1. Introduction

Engineering Asset Management as a discipline addresses valuable contributions of asset management to organization’s success [1]. Good asset management is becoming an expected practice in mature organizations all over the world. PAS 55:2008, which is the first publicly available specification for optimized management of physical assets, was developed by a consortium of 50 organizations from 15 different industry sectors in 10 countries. Given the popularity of PAS 55, after consultation with industry and professional bodies around the world, the specification was put forward in 2009 to the International Standards Organization as the basis for a new ISO standard for asset management. This was approved and the resulting ISO 55000 family of standards have been developed with 31 participating countries [2]. A EAM certificate provides recognized credibility in good practice and corporate governance, and a robust platform for developing further improvements.

A number of utility service providers have obtained asset management certificates through hiring well-known consultancy companies to perform auditing and performance assessment on EAM. However, many small and medium-sized enterprises (SMEs) cannot afford to hire renowned consultancy companies to guide them in obtaining required certificates and provide more professional suggestions to optimize asset management. Therefore, one purpose of the current research in EAM is to build an intelligent system so that it can automatically classify different performance levels of a particular company and then identify the most suitable practice in EAM for that company after benchmarking with information and performances given by other companies. For the other, PAS-55 only lists general guidelines in what elements are required to be accomplished so as to obtain the certificate in EAM. There is no specific method to evaluate the standard implementations and measure performance for managing assets. This paper aims to fill this gap and rank performance levels of asset management automatically and rapidly after conducting survey, instead of evaluating questionnaires, analyzing results, and ranking their performances with a tedious process. To achieve this goal, a novel Bayesian inference method for finding optimal parameters of support vector machines is proposed in this paper. The
major reason why support vector machines are adopted is that it has many unique advantages in solving small samples, nonlinear and high-dimensional pattern recognition problems [3–5]. Moreover, after optimizing support vector machines, this newly method not only has accomplished the requirements for ranking performance levels, but also can improve effectiveness and efficiency of analyzing and measuring performance levels in a simplified and low costing way. Moreover, predicted performance levels can be provided to other small and medium-sized enterprises (SMEs) and industries for benchmarking and proceed further survey and research.

The novelties of the proposed Bayesian inference method are summarized as follows. Firstly, a state space model is constructed to establish the relationship between parameters of support vector machines and guess cross-validation accuracy. Here, the guess cross-validation aims to alleviate the overfitting problem in the training process of support vector machines. Secondly, particle filter is introduced to iteratively obtain posterior probability density functions of parameters of support vector machines. According to our literature review, the particle filter for one-dimensional optimization [6], wind farm layout design [7], slurry pump diagnosis [8], bearing fault diagnosis [9], etc., have been reported. However, its use for optimization of support vector machines can be extended to optimize parameters of other supervised learning methods.

The rest of this paper is outlined as follows. Fundamental theories related to the proposed method are simply reviewed in Section 2. The novel Bayesian inference method for finding optimal parameters of support vector machines for multiclass classification problems is proposed in Section 3. Industrial survey data are analyzed in Section 4 to demonstrate the effectiveness of the proposed Bayesian inference method, and comparisons with some randomly selected parameters are conducted. Conclusions are drawn at last.

2. Fundamental theories related to the proposed method

2.1. Support vector machine for binary classification problems and its extension for multiclass classification problems

Support vector machine [3] is a popular supervised learning method for many binary classification problems. Its fundamental theory is introduced in the following. Given a training data set \( T = \{(y_i, z_i) | y_i \in \mathbb{R}^p, z_i \in \{-1, 1\}\}^{n}_{i=1} \). Here, \( y_i \) is a \( p \)-dimensional real vector, \( z_i \) is a binary label, which belongs to either \(-1\) or \(1\). In some cases, if the training data set is linearly separable, two hyperplanes can be used to separate the training data set. Moreover, it is required that no training data are located in the margin of the two hyperplanes. By considering the two points, the margin between the two hyperplanes should be maximized as much as possible. Therefore, solving the optimization problem provided by Eq. (1) is able to find a maximum-margin hyperplane for binary classification problems:

\[
\begin{align*}
\text{arg min}_{\omega, c} & \left\| \omega \right\|^2 + C \sum_{i=1}^{n} \xi_i \\
\text{subject to} & \; \; v_i (\omega \cdot y_i - c) \geq 1 - \xi_i, \; \; \xi_i \geq 0.
\end{align*}
\]

(2)

To find a solution to Eq. (2), a Lagrangian equation with Lagrange multipliers \( \alpha_i \) and \( \beta_i \) is constructed as follows:

\[
\begin{align*}
\text{arg min}_{\omega, c} & \left\{ \frac{1}{2} \| \omega \|^2 + C \sum_{i=1}^{n} \xi_i \right\} \\
\text{subject to} & \; \; \alpha_i (\omega \cdot y_i - c) - 1 - \xi_i - \sum_{j=1}^{n} \beta_j y_i y_j \geq 0, \; \; \alpha_i, \beta_j \geq 0.
\end{align*}
\]

(3)

The derivative of Eq. (3) with respect to \( \omega \) results in:

\[
\omega = \frac{1}{n} \sum_{i=1}^{n} \alpha_i y_i.
\]

(4)

The derivative of Eq. (3) with respect to \( c \) results in:

\[
\sum_{i=1}^{n} \alpha_i = 0.
\]

(5)

Then, after Eqs. (4) and (5) are substituted to Eq. (3), an alternative form of Eq. (3) is given as follows:

\[
\begin{align*}
\text{arg max}_{\alpha} & \; \; \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1, i \neq j}^{n} \alpha_i \alpha_j y_i y_j \cdot y_j \\
\text{subject to} & \; \; \sum_{i=1}^{n} \alpha_i y_i = 0, \; \; \sum_{i=1}^{n} \alpha_i = \mathbb{1}, \; \; \alpha_i \geq 0.
\end{align*}
\]

(6)

By solving Eq. (6), a linear decision function for binary classification problems is obtained as follows:

\[
f(y) = \text{sign} (\omega \cdot y - c) = \text{sign} \left( \sum_{i=1}^{n} \alpha_i y_i y_j \cdot y_j - \frac{1}{2} \sum_{i=1}^{n} \alpha_i y_i y_j \cdot y_j - c \right).
\]

(7)

where only a small number of \( \alpha_i \) are non-zero, and their corresponding training data are called support vectors.

Additionally, if the training data are linearly inseparable, kernel methods [10–12] should be used to map the training data to a high-dimensional space, in which it is possible to linearly separate the training data. Here, the kernel function should satisfy Mercer’s theorem. In practice, two kernel functions including polynomial and Gaussian radial kernel functions are popular. Compared with the polynomial kernel function, the Gaussian radial kernel function is preferable because it has less parameters and a good performance for handling non-linear classification problems. In this paper, the Gaussian radial kernel function is chosen. Therefore, Eq. (7) is modified as follows:

\[
f(y) = \text{sign} \left( \sum_{i=1}^{n} \alpha_i y_i \exp (-\gamma \|y_i - y_j\|^2) - c \right).
\]

(8)

where \( \gamma \) is the kernel parameter and \( \| \cdot \| \) is the modulus of the \( p \)-dimensional real vector.

The aforementioned theory related to support vector machine can only be used to classify binary problems. If a multiclass problem is required to be solved, it is necessary to reduce the multiclass problem to multiple binary classification problems, namely multiple support vector machines. In other words, multiple support vector machines are required to be built for multiclass classification problems. There are two popular strategies, namely one-against-all strategy and one-against-one strategy. For details, please refer to [3].

According to Eqs. (6) and (8), given the training data set, it is not difficult to conclude that the two parameters, including the kernel parameter and the error penalty constant \( C \), should be optimized so as to achieve good performances for predictions of unknown testing data set. In this paper, particle filter based Bayesian inference on optimal parameters of support vector
machines for multiclass classification problems is presented in Section 3. Prior to its formal presentation, the fundamental theory of particle filter is reviewed in Subsection 2.2.

2.2. Particle filter

In Bayesian statistical inference, particle filter or sequential Monte Carlo [13,14] is a particle Monte Carlo methodology to solve filtering problems. Its fundamental theory is introduced in the following. Given some measurements \( z_{1:k} \), if \( N_t \) random particles \( \{ x^i_{1:k} \}_{i=1}^{N_t} \) and their associated weights \( \{ \omega^i_{1:k} \}_{i=1}^{N_t} \) can be directly drawn from a true probability density function \( p(x_{1:k} | z_{1:k}) \), the true probability density function is expressed as follows:

\[
p(x_{1:k} | z_{1:k}) \approx \frac{1}{N_t} \sum_{i=1}^{N_t} \omega^i_{1:k} \delta(x_k - x^i_k),
\]

where \( \delta(\cdot) \) is the Dirac delta function and \( x_k \) is the state or parameter at iteration \( k \). However, in some cases, because it is intractable to directly random the particles from the true posterior density function, it is necessary to resort to an importance function \( q(x_{1:k} | z_{1:k}) \) to generate the random particles. Moreover, if the true posterior density function \( p(x_{1:k} | z_{1:k}) \) is proportional to an analytical function \( \pi(x_{1:k} | z_{1:k}) \), the weight \( \omega^i_{1:k} \) is proportional to [13,14]:

\[
\omega^i_{1:k} \propto \frac{\pi(x_{1:k} | z_{1:k})}{q(x_{1:k} | z_{1:k})}, \quad i = 1, 2, \ldots, N_t.
\]

Additionally, if \( q(x_{1:k} | z_{1:k}) = q(x_{1:k} | x_{1:k-1}, z_{1:k}) \times q(x_{1:k-1} | z_{1:k-1}) \) and \( p(x_{1:k} | z_{1:k}) \propto p(z_k | x_k) \times p(x_{1:k-1} | z_{1:k-1}) \) are considered, Eq. (10) can be iteratively updated as follows [13,14]:

\[
\omega^i_{k-1} \propto \frac{p(x_{1:k} | z_{1:k})}{q(x_{1:k} | z_{1:k})} \times \frac{p(x_{1:k-1} | z_{1:k-1})}{q(x_{1:k-1} | z_{1:k-1})} \times \frac{p(z_{1:k} | x_{1:k})}{q(z_{1:k} | x_{1:k})}, \quad i = 1, 2, \ldots, N_t.
\]

(11)

If \( q(x_{1:k} | z_{1:k}) = q(x_{1:k} | x_{1:k-1}, z_k) \) and weight normalization are considered, Eq. (11) is modified as:

\[
\omega^i_k = \omega^i_{k-1} \times \left( \frac{p(z_k | x_k) \times p(x_1 | x_{1:k-1}, z_k)}{q(z_k | x_k) \times p(x_1 | x_{1:k-1}, z_k)} \right), \quad i = 1, 2, \ldots, N_t.
\]

(12)

In many practical cases, to simplify Eq. (12), it is preferable to choose the importance function as a prior distribution, namely \( q(x_{1:k} | x_{1:k-1}, z_k) = p(x_{1:k} | x_{1:k-1}) \) [13,14]. Finally, Eq. (12) is simplified as:

\[
\omega^i_k = \omega^i_{k-1} \times \left( \sum_{i=1}^{N_t} \omega^i_{k-1} \times p(z_k | x_k) \right), \quad i = 1, 2, \ldots, N_t.
\]

(13)

According to the aforementioned theory of particle filter, it is clear to see that, given the measurements \( z_{1:k} \), the probability density function \( p(x_{1:k} | z_{1:k}) \) can be posteriorly estimated, which is able to infer an optimal state or parameter. In Section 3, based on particle filter, a novel Bayesian inference method for finding optimal parameters of support vector machines will be proposed.

3. Novel Bayesian inference on optimal parameters of support vector machines for multiclass classification problems

As aforementioned in Section 2.1, for the use of support vector machines for multiclass classification problems, such as evaluation of performance of EAM, the two parameters, including the kernel parameter \( \gamma \) and the error penalty constant \( C \), must be optimized to achieve both high training and testing accuracies. Moreover, to avoid the overfitting problem, a well-known procedure called cross-validation [15] should be employed. In \( K \)-fold cross-validation, the training data set is artificially divided to \( K \) subsets with an equal size. One subset is used as validation data to test support vector machines, which are trained by the remaining \( K-1 \) subsets. Circularly, \( K \) subsets can be respectively used as validation data. Cross-validation accuracy (CVA) is defined as the percentage of validation data which are correctly classified. Consequently, in this paper, finding optimal parameters of support vector machines becomes a two-dimensional optimization problem, which aims to maximize cross-validation accuracy. To posteriorly infer optimal parameters of support vector machines, the following state space model is constructed:

\[
\gamma_k = \gamma_{k-1},
\]

(14)

\[
C_k = C_{k-1}.
\]

(15)

\[
\text{CVA}_k = \text{CVA}(\gamma_k, C_k) - m_k,
\]

(16)

where \( m_k \) is the measurement noise and it is subject to an uniform distribution U(0,100). Here, the range of the uniform distribution is from 0 to 100 because CVA is always limited to a range from 0% to 100%. CVA is the abbreviation of guess cross-validation accuracy, which will be constructed later. According to Bayes’ theorem, the posterior probability density function \( p(\gamma_k, C_k | \text{CVA}_{1:k}) \) can be inferred from Eqs. (14) to (16).

Using Bayes’ theorem, the joint posterior density function \( p(\gamma_k, C_k | \text{CVA}_{1:k}) \) is represented by:

\[
p(\gamma_k, C_k | \text{CVA}_{1:k}) = \frac{p(\text{CVA}_{1:k} | \gamma_k, C_k) p(\gamma_k, C_k | \text{CVA}_{1:k-1})}{p(\text{CVA}_{1:k} | \text{CVA}_{1:k-1})} = \frac{p(\gamma_k, C_k | \text{CVA}_{1:k-1})}{p(\text{CVA}_{1:k} | \text{CVA}_{1:k-1})}
\]

(17)

It is interesting to find that \( p(\gamma_k, C_k | \text{CVA}_{1:k}) \) can be iteratively calculated according to Eq. (17). Moreover, if CVA is monotonically increasing, optimal parameters of support vector machines can be posteriorly inferred from Eq. (17) by considering the following inequality:

\[
0 \leq \text{CVA}_1 \leq \text{CVA}_2 \leq \ldots \leq \text{CVA}_K \leq \text{CVA}(\gamma_{\text{optimal}}, C_{\text{optimal}}) \leq 100.
\]

(18)

The details for realizing Eqs. (17) and (18) by using particle filter are clarified as follows.

Step 1. \( N_t = 1000 \) random particles \( \{ \gamma^i_0, C^i_0 \}_{i=1}^{N_t} \) with equal weights \( \{ \omega^i_0 \}_{i=1}^{N_t} = 1/N_t \) are drawn from an uniform distribution. Here, \( \gamma \) is limited to a range from 2 to 20. \( C \) is limited to a range from 2 to 25. According to the theory of cross-validation, these random particles result in 1000 cross-validation accuracies. Then, arrange these cross-validation accuracies from the lowest value to the highest value. The first guess cross-validation accuracy GCVA1 is obtained by taking the 50th percentile of the 1000 cross-validation accuracies. Then, the posterior probability density function of \( \gamma_1, C_1 \) (GCVA1) is derived as follows:

\[
p(\gamma_1, C_1 | \text{GCVA}_1) = \frac{\sum_{i=1}^{N_t} \omega^i_0 \times \delta\left( (\gamma_1, C_1) - (\gamma^i_0, C^i_0) \right)}{\sum_{i=1}^{N_t} \omega^i_0 \times \delta\left( (\gamma_1, C_1) - (\gamma^i_0, C^i_0) \right)}
\]

(19)

where \( \omega^i_1 \) is calculated by:

\[
\omega^i_1 = \left\{ \begin{array}{ll}
\omega^i_0 \times \frac{1}{N_t} \text{CVA}(\gamma^i_0, C^i_0) \geq \text{CVA}_1, \\
0 & \text{CVA}(\gamma^i_0, C^i_0) < \text{CVA}_1,
\end{array} \right.
\]

(20)
According to Eq. (20), if the cross-validation accuracies \( CVA(\gamma^i_k, C^i_k) \) are smaller than the first guess cross-validation accuracy, their associated weights are set to zero. In other words, parts of the random particles are discarded in terms of their corresponding low cross-validation accuracies. It should be noted that if a few iterations are conducted, most of the weights will become negligible and the variances of the weights increase. This problem is called a degeneracy problem in particle filter. To relieve this problem, a systematical resampling method should be conducted. Firstly of all, cumulate all weights by using \( c_1 = 0 \) and \( c_i = c_{i-1} + \omega^i_k \), and form a cumulative function. Draw a value \( a_i \) from an uniform distribution \( U[0, N_i^{-1}] \).

Then, for \( a_i = a_1 + N^{-1}(j-1), j = 1, 2, \ldots, N_i \) along the cumulative function, if \( a_i \geq c_i \) is satisfied, \( i \leftarrow i + 1 \). Then, \( p^i = p^i \) and \( C^i = C^i \). At last, all weights are set to \( 1/N_i \).

Step 2. Draw \( N_t \) random particles \( \{ \gamma^i_k, \sigma^i_k \}_{i=1}^{N_t} \) with equal weights \( \{ \omega^i_k \}_{i=1}^{N_t} = 1/N_t \) from the posterior distribution \( p(\gamma^i_k, C^i_k | CVA^i_{GCV}) \). According to the theory of cross-validation, these random particles result in 1000 cross-validation accuracies. Arrange these cross-validation accuracies from the lowest value to the highest value. The kth guess cross-validation accuracy \( CVA_k \) is obtained by taking the 50th percentile of the 1000 cross-validation accuracies. To keep the kth guess cross-validation accuracy monotonically increasing, the kth guess cross-validation accuracy must be larger than all of the previous guess cross-validation accuracies. If not, the recent largest guess cross-validation accuracy should be employed as the kth guess cross-validation accuracy. The posterior probability density function \( p(\gamma_k, C_k | CVA_{GCV}) \) is estimated by:

\[
p(\gamma_k, C_k | CVA_{GCV}) = \frac{\sum_{i=1}^{N_t} \omega^i_k \times \delta(\gamma_k, C_k - \{ \gamma^i_k, \sigma^i_k \})}{\sum_{i=1}^{N_t} \omega^i_k \times \delta(\gamma_k, C_k - \{ \gamma^i_k, \sigma^i_k \}) d\gamma_k dC_k} = \frac{\sum_{i=1}^{N_t} \omega^i_k \times \delta(\gamma_k, C_k - \{ \gamma^i_k, \sigma^i_k \})}{\sum_{i=1}^{N_t} \omega^i_k}.
\]  
(21)

where \( \omega^i_k \) is calculated by:

\[
\omega^i_k = \left\{ \begin{array}{ll}
\frac{CVA(\gamma^i_k, C^i_k) \geq CVA_k}{\sum_{i=1}^{N_t} \omega^i_k}, & i = 1, 2, \ldots, N_t; \\
0, & CVA(\gamma^i_k, C^i_k) < CVA_k.
\end{array} \right.
\]  
(22)

Accordingly, if the cross-validation accuracies \( CVA(\gamma^i_k, C^i_k) \) are smaller than the kth guess cross-validation accuracy \( CVA_k \), their associated weights are set to zero. Resample \( N_t \) random particles \( \{ \gamma^i_k, \sigma^i_k \}_{i=1}^{N_t} \) from Eq. (21) by the aforementioned systematic resampling method.

Step 3. Increase \( k = k + 1 \) and repeat Step 2 until k exceeds a specified maximum iteration number. It should be noted that calculation time of the proposed Bayesian inference method becomes extensive as the specified maximum iteration number increases and the number of random particles. Therefore, the specified maximum iteration number should not be set to a large value and is equal to 4 in this paper so as to reduce calculation time. Because the posterior probability density function \( p(\gamma_k, C_k | CVA_{GCV}) \) is estimated, optimal parameters of support vector machines can be derived by taking the random particle corresponding to the Lth percentile of all the cross-validation accuracies generated from the posterior probability density function \( p(\gamma_k, C_k | CVA_{GCV}) \). In this paper, \( L \) is set to 100.

4. A case study in Hong Kong

In Hong Kong (HK), certificates related to EAM have been awarded to a number of public utilities corporations, such as China Light and Power Co. Ltd. (CLP), Mass Transit Railway Corporation (MTRC), the Hong Kong and China Gas Co. Ltd. (TG), etc. Some E&M buildings, small and medium-sized enterprises (SMEs), services organizations, and plants as a substantial part of EAM have not adopted the EAM standard completely. The consultancy fee for accomplishing asset management certificate is extremely high, and no specialized department or organization evaluates performance levels for asset management in Hong Kong. In order to investigate the performance of asset management and EAM standard applications, sampling survey was conducted. According to the PAS-55 guideline, the structured questionnaire was designed, and all questionnaires were sent to 40 Operation and Maintenance (OM) departments/companies in Hong Kong, the objects of investigation included: public services of HK government, commercial, residential, industrial and composite buildings portfolio.

For our questionnaire design, it included general information about companies, information management standards, and significance levels for standard implementation. There were 60 questions in total and the main part focused on guidelines for information management in EAM. General information included categories of operations of respondents, the number of employees, number of years in business operation, and total operation and maintenance cost of buildings/plans. Then, the questionnaire listed criteria of asset information management which needed to be applied. Besides filling questionnaire, face-to-face interviews were conducted by a researcher, which could explain more about the purpose and specific meaning for respondents. It improved accuracy and efficiency of survey data. For another, some questions were required to be discussed in order to obtain the comprehensive information. After interviewers finished questionnaire, the researcher had to check if anything was missed or overlooked. During the whole process, the interviewers could communicate with the researcher and ask for further explanations to the listed questions. Consequently, the integrity and validity check were

**Fig. 1.** The flowchart of the proposed Bayesian inference method for finding optimal parameters of support vector machines for multiclass classification problems.
completed for all questionnaires. For the details of questionnaires, please see the Appendix.

According to information management standard in PAS-55, it lists 53 criteria which belong to four parts. Since 40 departments/organizations stand for 40 samples, and there are 53 attributes for each sample. Therefore, collected survey data size is 40 samples with 53 attributes, and their associated performance levels are 1, 2 and 3. The performance levels (labels of support vector machines) are defined as follows. Performance level 1 has a satisfied performance for standard adoption with a high percentage in achievements which is ranged from 75% to 100%. Performance level 2 shows an average performance compared with performance level 1 and performance level 3. The percentage of standards adoption is from 45% to 74%. Performance level 3 stands for companies/organizations have a low percentage in standard application, which is in a range of 20% to 44%. They may ignore some critical criteria, or they do not comply with the standard step by step. Hence, a few potential risks should be emphasized and some processes are required to be immediately improved.

Half of the samples are regarded as the training data set. The rest of them are treated as the testing data set. Because the sample size of the training data set is only 20, 5-fold cross-validation is used in the proposed Bayesian inference method for finding optimal parameters of support vector machines. The optimal parameters found by the proposed Bayesian inference method are listed in Table 1. The support vector machines with the optimal parameters are used to predict the training data set and result in the high training accuracy of 100%. Moreover, the support vector machines with the same optimal parameters are used to predict the unknown testing data set and generate the high testing accuracy of 95%. To highlight the superiority of the proposed Bayesian inference method, some randomly selected parameters are used in support vector machines for processing the same training and testing data sets. The results are tabulated in Table 1, where it is clear to find that the randomly selected parameters cause the overfitting problem of support vector machines. Moreover, we observed that the higher the parameter $C$ is, the higher the training accuracy is; the lower the parameter $\gamma$ is; the higher the testing accuracy is. Nevertheless, for optimization of support vector machines and achieving both high training and testing accuracies, the parameters $C$ and $\gamma$ should be jointly optimized.

In this paper, all codes were made based on the software called ‘MATLAB’. To realize support vector machines, we used the famous MATLAB software pack ‘LIBSVM’ developed by Chang and Lin [16]. To realize Bayesian inference on optimal parameters of support vector machines, we developed the MATLAB codes by ourselves. The calculation time of the proposed method was measured by the MATLAB software pack installed in a computer with CPU 2.33 GHz and 2 G RAM and it was equal to 15 s. The calculation time was not so fast because it was affected by the number of the random particles and the computer performance.

5. Discussion

The final testing results shows that the 19 out of 20 samples have been correctly classified. The 5% misclassification was possibly caused by the samples, which were collected from the different business natures, and the performance requirements and critical business may not be completely consistent. What is more, there are not huge amount data available to support training and testing data samples, since asset management department is seldom set up in the companies currently. For the further applications of our proposed method, after more companies fill the questionnaire, the proposed method will automatically classify companies’ performance. According to classification results, respondents will be aware of their current performance and improve better referring to the engineering asset management guideline and specifications. Moreover, through investigations from different companies and organizations, data sample sizes will be increased and updated continuously, and prediction accuracy will be improved as well.

6. Conclusion

In this paper, the intelligent classification method for multiclass classification problems, such as evaluation of performance of EAM, was developed by using the optimized support vector machines. To find optimal parameters of support vector machines, a novel Bayesian inference method was proposed. Firstly, the state space model was constructed to establish the relationship between the parameters of support vector machines, including the kernel parameter $\gamma$ and the error penalty constant $C$, and the guess cross-validation accuracies. The guess cross-validation accuracies were monotonically increasing and aimed to alleviate the overfitting problem of support vector machines. Secondly, the particle filter was introduced to use an amount of the random particles and their associated weights to posteriorly infer the probability density functions of the parameters of support vector machines. Once their posterior probability density functions were derived, the optimal parameters of support vector machines could be found. The industrial survey data were investigated to illustrate how the proposed Bayesian inference method worked. The results showed that the proposed Bayesian inference method is able to produce the high training accuracy of 100% and the high testing accuracy of 95%, while the other randomly selected parameters can only produce the high training accuracies but the low testing accuracies. In other words, if the parameters of support vector machines for evaluating performance of EAM were incorrectly chosen, the overfitting problem happened.

In this paper, we provided a relatively new idea, namely Bayesian inference, and an alternative way to optimize support vector machines. In our future work, we will theoretically and thoroughly compare our proposed method with other optimization methods, such as genetic algorithm [17], particle swarm optimization [18], etc. Furthermore, we will consider to apply the proposed Bayesian inference to optimize parameters of other supervised learning methods because support vector machine is just one kind of supervised learning methods.

Acknowledgement

This work described in this paper is partly supported by a grant from National Natural Science Foundation of China (Project No. 51505307), a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Nos. CityU_122513) and a grant from City University of Hong Kong (Project No. 7004251). The authors would like to thank the valuable comments provided by reviewers. Additionally, the first author would like to express her sincere appreciation to Dr. Dong Wang's comments and suggestions.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Parameters of support vector machines and training and testing accuracies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The optimal parameters obtained by the proposed method</strong></td>
<td><strong>Some randomly selected parameters</strong></td>
</tr>
<tr>
<td>$C=3.1591 \times 10^4$, $\gamma=3.7764 \times 10^{-4}$</td>
<td>$C=1$, $\gamma=10^{-3}$</td>
</tr>
<tr>
<td>Training accuracy 100%</td>
<td>45%</td>
</tr>
<tr>
<td>Testing accuracy 95%</td>
<td>35%</td>
</tr>
<tr>
<td></td>
<td>$C=1 \times 10^3$, $\gamma=10^{-1}$</td>
</tr>
<tr>
<td></td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>$C=1 \times 10^2$, $\gamma=10$</td>
</tr>
<tr>
<td></td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>$C=1 \times 10^4$, $\gamma=10$</td>
</tr>
<tr>
<td></td>
<td>100%</td>
</tr>
</tbody>
</table>
Appendix

Questionnaire

Scoring Scheme

For each question, please indicate the degree to which the auditor agrees or disagrees with the statements by putting a cross, i.e. “X”, in the space provided which is corresponding to one of the following 5 scales and 2 other categories, namely:

<table>
<thead>
<tr>
<th>Totally Adopted (100-91%)</th>
<th>Mostly Adopted (90%-75%)</th>
<th>Generally Adopted (74-41%)</th>
<th>Slightly Adopted (40-11%)</th>
<th>Not Adopted (10-0%)</th>
<th>More than those Adopted</th>
<th>Neutral / No Need</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

**Questionnaires (Theme on Focus Pattern of Information Management of EAM)**

4.4.6 Information Management of BSI PAS 55 – Engineering Asset Management (EAM)

4.4.6(a) Adequacy of Information Authorized for Use of Asset Management

<table>
<thead>
<tr>
<th>4.4.6(a) Adequacy of Information Authorized for Use of Asset Management</th>
<th>Adopted</th>
<th>Not Adopted</th>
<th>Special</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

**General**

1. Information Manager assigned by the Management Representative Specifically

2. Information Structure established by Information Manager


4. Level control of information accuracy to deliver Asset Management Strategy, Objectives and Plans
<table>
<thead>
<tr>
<th><strong>Completeness of Information to deliver Asset Management Strategy, Objectives and Plans</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>5. Information to enable optimization and prioritization of Asset Management Strategy &amp; Plans</td>
</tr>
<tr>
<td>6. Information to assess financial benefits of Asset Management Plans of Improvements</td>
</tr>
<tr>
<td>7. Information to determine operational and financial impact on unavailability or failure of the major operations</td>
</tr>
<tr>
<td>8. Information to compare life cycle costs among alternative investments of the major engineering asset</td>
</tr>
<tr>
<td>9. Information to monitor details and expiry dates of licenses, warranties and certifications, etc.</td>
</tr>
<tr>
<td>10. Information provided to determine with costs of activities and replacements with track record of market prices</td>
</tr>
<tr>
<td>11. Information to determine end of economic life of the major engineering asset with track records of paid rates</td>
</tr>
<tr>
<td>12. Information allowed for performing financial analysis of planned income and expenditures</td>
</tr>
<tr>
<td>13. Information to determine financial and resource impact on availability and performance over a contingency period if contingency plan is taken place</td>
</tr>
<tr>
<td>14. Information to assess overall financial performance of the engineering assets</td>
</tr>
<tr>
<td>15. Information allowed to perform risk analysis for operation and maintenance works</td>
</tr>
<tr>
<td>16. Information to assure performance of statutory compliance with track records with respect to the rules</td>
</tr>
</tbody>
</table>

4.4.6(b) Periodic Review and Revision to Maintain Adequacy of the Information Management System

<p>| 4.4.6(b) Periodic Review and Revision to Maintain | Adopted | Not | Special |</p>
<table>
<thead>
<tr>
<th>Adequacy of the Information Management System</th>
<th>Adopted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>17. Consistent coded names of asset items information made available for identification and definition of systems</td>
<td>☐</td>
</tr>
<tr>
<td>18. Information to manage asset life cycles their legal and regulatory management requirements</td>
<td>☐</td>
</tr>
<tr>
<td>19. Information to describe assets, functions and systems being served</td>
<td>☐</td>
</tr>
<tr>
<td>20. Information to give unique asset identification and asset registers</td>
<td>☐</td>
</tr>
<tr>
<td>21. Information to give locations and spatial layout of assets</td>
<td>☐</td>
</tr>
<tr>
<td>22. Information to give engineering data, design parameters, and drawings</td>
<td>☐</td>
</tr>
<tr>
<td>23. Information to give vendor data for assets</td>
<td>☐</td>
</tr>
<tr>
<td>24. Information to give testing and commissioning dates and data of assets</td>
<td>☐</td>
</tr>
<tr>
<td>25. Information to access planning and work O&amp;M schedules</td>
<td>☐</td>
</tr>
<tr>
<td>26. Information to give task risk assessments and control measures</td>
<td>☐</td>
</tr>
<tr>
<td>27. Information to give task details of the last maintained / inspected and when they are next due.</td>
<td>☐</td>
</tr>
<tr>
<td>28. Information to give listing of overdue / outstanding tasks</td>
<td>☐</td>
</tr>
<tr>
<td>29. Information to give historical record of planned and unplanned maintenance tasks performed</td>
<td>☐</td>
</tr>
<tr>
<td>30. Information to give operational data including performance characteristics and design limits</td>
<td>☐</td>
</tr>
<tr>
<td>31. Information to give financial data of available cost, cost of historical pm tasks, operating cost, downtime impact, replacement value, initial cost, etc.</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td>Information to give working programmes and schedules of works and settings (long and short terms)</td>
</tr>
<tr>
<td>---</td>
<td>-------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>32.</td>
<td></td>
</tr>
<tr>
<td>33.</td>
<td>Information to give planning of asset possession, shutdown and outage.</td>
</tr>
<tr>
<td>34.</td>
<td>Information to give operating details of condition monitoring systems</td>
</tr>
</tbody>
</table>

**4.4.6(c) Allocation of Appropriate Roles and Responsibilities and Authorities in using the Information Management System**

<table>
<thead>
<tr>
<th></th>
<th>4.4.6(c) Allocation of Appropriate Roles and Responsibilities and Authorities in using the Information Management System</th>
<th>Adopted</th>
<th>Not Adopted</th>
<th>Special</th>
</tr>
</thead>
<tbody>
<tr>
<td>35.</td>
<td>Information accessible and available to all relevant personnel under monitoring and controls</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36.</td>
<td>Allocation of responsibilities and authorities for maintenance, access, archiving and disposal of information.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37.</td>
<td>Information maintenance, version control and assurance activities.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38.</td>
<td>Information generation, capture or importing of the identified items</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39.</td>
<td>Information ownership and maintenance demarcation where assets interface across a system or network of assets</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40.</td>
<td>Information of asset build-up conditions and duty use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41.</td>
<td>Asset service requirements, conditions, and performance targets or standards</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42.</td>
<td>Requirements of key performance indicators</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43.</td>
<td>Information of current tasks and planned works</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44.</td>
<td>Criteria of non-conformance and the actions to be taken</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Details of emergency plans, responsibilities and contacts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---------------------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>45.</td>
<td>Information of materials, inventory, purchasing management systems</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46.</td>
<td>Information of decision-supporting systems for optimization and life cycle costing models</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47.</td>
<td>Information of service performance reporting systems</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48.</td>
<td>Information of staff locations, scheduling and dispatch systems</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49.</td>
<td>Information of capital expenditure planning and condition monitoring systems</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.4.6(d) Assurance against Unintended Use of Obsolete Information in using the Information Management System (assessment under a separate cover)

4.4.6(e) Assurance of Archival Information retained for Legal or Knowledge Preservation in the Information Management System (assessment under a separate cover)

**4.4.6(f) Assurance of Information Security with Back-up Recovery in the Information Management System**

<table>
<thead>
<tr>
<th></th>
<th>Adopted</th>
<th>Not Adopted</th>
<th>Special</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.4.6(f) Assurance of Information Security with Back-up Recovery in the Information Management System</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51. Storing information items according to integrity, security and confidentiality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52. Performing management cycles of establishment, implementation, retention, and disposal of records</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53. Monitoring effectiveness of record procedures, access controls and storage facilities and disposal</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Conclusions**

(A) **What Damages if Any if Part of the Information Management System of BSI PAS-55 Not in Use for the O&M of the Engineering Asset**

<table>
<thead>
<tr>
<th></th>
<th>Significant</th>
<th>Not Significant</th>
<th>Special</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| What if the Information Management of BSI PAS-55 not in use for the O&M of the Engineering Asset |   |   |   |   |   |   |
### 54. Damages if any when PAS-55 Clause 4.4.6(a)
Information Authorization for Use not adopted

Main Reason for this answer if any: __________

### 55. Damages if any when PAS-55 Clause 4.4.6(b)
Periodic Review on Revision to Maintain Use not adopted

Main Reason for this answer if any: __________

### 56. Damages if any when PAS-55 Clause 4.4.6(c)
Allocation of Roles and Responsibility on Use not adopted

Main Reason for this answer if any: __________

### 57. Damages if any when PAS-55 Clause 4.4.6(f)
Assurance and Back-up Recovery on Use not adopted

Main Reason for this answer if any: __________

### (B) Overall Management Performance of the O&M of the Engineering Asset

<table>
<thead>
<tr>
<th>Satisfaction Level of the Present Operation and Maintenance Performance</th>
<th>Satisfactory --- Not Satisfactory</th>
<th>Special</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

#### 58. Good Practice in Cost Engineering

Main Reason for this answer if any: __________

#### 59. Good Practice in Quality / Reliability Management

Main Reason for this answer if any: __________
References


Jingjing Zhong is a PhD candidate in the Department of Systems Engineering & Engineering Management, City University of Hong Kong. She received her Master's in Management, Economics and Industrial Engineering from Politecnico di Milano in Italy in 2009. Her research interests include Engineering Asset Management standards PAS-55, performance management, benchmarking and maintenance management.

Dr. Peter W. Tse is currently the Group Leader of the Smart Engineering Asset Management Laboratory (SEAM) and the Director of Croucher Optical Non-Destructive Testing Laboratory in the Department of Systems Engineering and Engineering Management at the City University of Hong Kong (CityU). SEAM was established through generous donations from the industry of Hong Kong. The mission of SEAM is to provide support to industry for achieving near-zero breakdown of equipment and maintaining high quality services through the smart management of assets. As of today, SEAM has research collaboration/consultancy projects with over 30 international and local companies. Dr. Tse is the O-Committee Member of the Technical Committees of Non-Destructive Testing (TC 199), Safety of Machinery (TC 135) and Mechanical Vibration and Shock (TC 108) of the International Organization for Standardization (ISO). Currently he is a registered Professional Engineer in Canada, a Chartered Engineer in United Kingdom. He has been awarded the PCN (Personnel Certification in Non-Destructive Testing) Certificate of Competence in Condition Monitoring from the British Institute in Non-Destructive Testing (BINDT) and completed the Vibration Training Course - Vibration Analysis Level 2 in according with ISO 18436 Part 2. As of today, he has published over 250 articles in various international journals, proceedings, and professional reports.