Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

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

ARTICLE INFO

Article history:
Received 1 September 2015
Received in revised form
28 December 2015
Accepted 28 December 2015

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

ABSTRACT

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

* Corresponding author.
E-mail addresses: jingzhong2-c@my.cityu.edu.hk (J. Zhong), meptse@cityu.edu.hk (P.W. Tse), dongwang4-c@my.cityu.edu.hk (D. Wang).

http://dx.doi.org/10.1016/j.neucom.2015.12.132
0925-2312/

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 prognosis [8], bearing fault diagnosis [9], etc., have been reported. However, its use for optimization of supervised learning methods, particularly support vector machines, is very limited and seldom reported. The contents reported in this paper could be used to clarify how the particle filter is able to find optimal parameters of support vector machines. Moreover, because support vector machine is just one kind of supervised learning methods, the proposed Bayesian inference on optimal parameters 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 Bayesian inference 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 $\mathrm{T}=\left\{(\mathbf{y}_i, z_i)\,|\,\mathbf{y}_i \in \mathrm{R}^p, v_i \in \{-1,1\}\right\}_{i=1}^{n}$. Here, $\mathbf{y}_i$ is a p-dimensional real vector. v_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:

$$\begin{aligned}&\arg\min \tfrac{1}{2}\|\boldsymbol{\omega}\|^2\\&\text{subject to}\quad v_i(\boldsymbol{\omega}\cdot\mathbf{y}_i-c)\ge 1\,,\end{aligned}\tag{1}$$

where $\boldsymbol{\omega}$ is a normal vector to the hyperplane. $\cdot$ is the dot product operator and c is an offset of the hyperplane from the origin along the normal vector.

In some cases, if the constraints used in Eq. (1) are not satisfied, a slack variables ξ_i and an error penalty constant C should be used to consider the tradeoff between a large margin and an error penalty. Then, Eq. (1) is revised as:

$$\begin{aligned}&\arg\min_{\boldsymbol{\omega},\xi}\tfrac{1}{2}\|\boldsymbol{\omega}\|^2+C\sum_{i=1}^{n}\xi_i\\&\text{subject to}\quad v_i(\boldsymbol{\omega}\cdot\mathbf{y}_i-c)\ge 1-\xi_i,\quad \xi\ge 0\,.\end{aligned}\tag{2}$$

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

$$\arg\min_{\boldsymbol{\omega},\xi,\alpha,\beta,b}\frac{1}{2}\|\boldsymbol{\omega}\|^2+C\sum_{i=1}^{n}\xi_i-\sum_{i=1}^{n}\alpha_i\left[v_i(\boldsymbol{\omega}\cdot\mathbf{y}_i-c)-1+\xi_i\right]-\sum_{i=1}^{n}\beta_i\xi_i,\quad \alpha_i,\beta_i\ge 0\,.\tag{3}$$

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

$$\boldsymbol{\omega}=\sum_{i=1}^{n}\alpha_i v_i\mathbf{y}_i.\tag{4}$$

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

$$\sum_{i=1}^{n}\alpha_i v_i=0.\tag{5}$$

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

$$\begin{aligned}&\arg\max_{\alpha}\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1,}^{n}\alpha_i\alpha_j v_i v_j\mathbf{y}_i\cdot\mathbf{y}_i\ \text{subject to}\quad C\ge\alpha_i\ge 0\,,\\&i=1,\dots,n,\quad \sum_{i=1,}^{n}\alpha_i v_i=0.\end{aligned}\tag{6}$$

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

$$f(\mathbf{y})=\mathrm{sign}(\boldsymbol{\omega}\cdot\mathbf{y}-c)=\mathrm{sign}\left(\left(\sum_{i=1}^{n}\alpha_i v_i\mathbf{y}_i\right)\cdot\mathbf{y}-c\right)=\mathrm{sign}\left(\sum_{i=1}^{n}\alpha_i v_i\mathbf{y}_i\cdot\mathbf{y}-c\right),\tag{7}$$

where only a small number of α_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:

$$f(\mathbf{y})=\mathrm{sign}\left(\sum_{i=1}^{n}\alpha_i v_i\exp\left(-\gamma\|\mathbf{y}_i-\mathbf{y}\|^2\right)-c\right),\tag{8}$$

where γ 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 parameterγ 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

Please cite this article as: J. Zhong, et al., Novel Bayesian inference on optimal parameters of support vector machines and its application to industrial survey data classification, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2015.12.132