Expert Systems With Applications 59 (2016) 217–225

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Sliding window-based support vector regression for predicting micrometeorological data

Yukimasa Kaneda[a,*], Hiroshi Mineno[b,c]

[a] Graduate School of Integrated Science and Technology, Shizuoka University, 3-5-1 Johoku, Naka-ku, Hamamatsu, Shizuoka 432-8011, Japan
[b] College of Informatics, Academic Institute, Shizuoka University, 3-5-1 Johoku, Naka-ku, Hamamatsu, Shizuoka 432-8011, Japan
[c] JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan

ARTICLE INFO

Article history:
Received 4 February 2016
Revised 29 March 2016
Accepted 13 April 2016
Available online 23 April 2016

Keywords:
Predicting micrometeorological data
Data extraction
Dynamic aggregation
Support vector regression
Ensemble learning

ABSTRACT

Sensor network technology is becoming more widespread and sophisticated, and devices with many sensors, such as smartphones and sensor nodes, have been used extensively. Since these devices have more easily accumulated various kinds of micrometeorological data, such as temperature, humidity, and wind speed, an enormous amount of micrometeorological data has been accumulated. In recent years, it has been expected that such an enormous amount of data, called big data, will produce novel knowledge and value. Accordingly, many current applications have used data mining technology or machine learning to exploit big data. However, micrometeorological data has a complicated correlation among different features, and its characteristics change variously with time. Therefore, it is difficult to predict micrometeorological data accurately with low computational complexity even if state-of-the-art machine learning algorithms are used. In this paper, we propose a new methodology for predicting micrometeorological data, sliding window-based support vector regression (SW-SVR) that involves a novel combination of support vector regression (SVR) and ensemble learning. To represent complicated micrometeorological data easily, SW-SVR builds several SVRs specialized for each representative data group in various natural environments, such as different seasons and climates, and changes weights to aggregate the SVRs dynamically depending on the characteristics of test data. In our experiment, we predicted the temperature after 1 h and 6 h by using large-scale micrometeorological data in Tokyo. As a result, regardless of testing periods, training periods, and prediction horizons, the prediction performance of SW-SVR was always greater than or equal to other general methods such as SVR, random forest, and gradient boosting. At the same time, SW-SVR reduced the building time remarkably compared with those of complicated models that have high prediction performance.

1. Introduction

Sensor network technology is becoming more widespread and sophisticated, and devices with many sensors have been used extensively. The devices can very easily obtain various kinds of micrometeorological data such as temperature, humidity, and wind speed. Micrometeorological data is affected strongly by the surface of the earth and is related to our lives and industrial activity. Accordingly, the data has been used by many applications such as environmental control systems for greenhouses (Othman & Shazali, 2012; Park & Park, 2011). Moreover, more advanced applications exploit the data to a greater extent by using machine learning and data mining technology. Furthermore, an enormous amount of micrometeorological data has been accumulated by many devices, and it has been expected that analyzing such an enormous amount of data, called big data, will produce novel knowledge and value.

To predict micrometeorological data effectively, a number of researchers have studied machine learning (Smith, Hoogenboom, & McClendon, 2009). These researchers described prediction methods for micrometeorological data; particularly, prediction performance and computational complexity were often mentioned. Meanwhile, micrometeorological data has a complex correlation among different features such as temperature and humidity. Moreover, its characteristics change variously with time. Therefore, even if big data is given as training data, it is not easy to predict micrometeorological data accurately. Furthermore, in many cases, so that models can have high prediction performance, they have to become complicated, and the computational complexity increases. Accordingly, some models probably cannot be built from big data in a

* Corresponding author.
E-mail address: kaneda@minelab.jp (Y. Kaneda).

http://dx.doi.org/10.1016/j.eswa.2016.04.012

practical amount of computing time. In other words, there is a trade-off relationship between high prediction performance and low computational complexity. However, compatibility is required in some practical use. As the prediction performance in applications becomes higher, the quality provided by the applications becomes better. For example, in the case of environmental control systems based on prediction (Kolokotsa, Pouliezos, Stavrakakis, & Lazos, 2009), the higher prediction performance enables the systems to provide precise control, precise management, and better environments. On the other hand, models that need a long time for training are worthless in practical use. In current situations where the amount of usable data has increased remarkably, this trade-off relationship has become a more critical issue.

Recently, one type of machine learning algorithm, support vector machines (SVMs), have been used successfully in various fields. The basic theory is a more efficient learning method based on probably approximately correct (PAC) learning. Moreover, SVMs can separate non-linear data with low computational complexity. Since most data observed in the real world is likely to have non-linear relationships, SVMs have also been applied to micrometeorological data prediction (Antonanzas, Urraca, Martinez-de-Pison, & Antonanzas-Torres, 2015; Mohammadi, Shamshirband, Anisi, Alam, & Petković, 2015; Urraca, Antonanzas, Martinez-de-Pison, & Antonanzas-Torres, 2015). Moreover, SVMs led to better prediction performance than other algorithms such as artificial neural networks (ANNs) and the autoregressive integrated moving average (ARIMA) model (Chevalier, Hoogenboom, McClendon, & Paz, 2011; Maity, Bhagwat, & Bhatnagar, 2010). However, when SVMs learn big data, the computational complexity is still a matter of concern. Another alternative learning method, ensemble learning, has also been used more widely for predicting micrometeorological data (Singh, Gupta, & Rai, 2013). The prediction performance of ensemble learning is greater than or equal to that of SVMs. The basic methodology is a combination of weak learners built from different kinds of training data. The combination yields a higher generalizing capability that a single model cannot represent. In particular, some researchers proposed improved methods that could be applied to micrometeorological data prediction (Wang & Japkowicz, 2009; Xie, Li, Ngai, & Ying, 2009). However, it is difficult to apply the methods to regression, and it is possible that the models will not be able to follow micrometeorological data whose characteristics always change with time.

In this paper, we propose a new methodology for predicting micrometeorological data, sliding window-based support vector regression (SW-SVR). SW-SVR involves a novel combination of support vector regression (SVR) and ensemble learning. To represent complicated micrometeorological data easily, SW-SVR builds several SVRs specialized for each representative data group in various natural environments, such as different seasons and climates. The specialized SVRs are built based on our previous proposed method, dynamic short-distance data collection (D-SDC) that extracts effective data for specific data prediction by taking account of movements: changes in data during prediction horizons. Each weak learner built from each extracted data specializes on specific data and predicts accurately the data similar to the specialized data. Then, SW-SVR aggregates all the predicted values based on weights decided by the similarity between test data and each data specialized by weak learners. This new ensemble learning methodology that changes weights dynamically enables following micrometeorological data whose characteristics hardly change with time. Our results demonstrated that the prediction performance of SW-SVR was always greater than or equal to that of other general methods such as SVR, random forest, and gradient boosting. At the same time, SW-SVR reduced the building time remarkably compared with that of complicated models that have high prediction performance.

2. Related work

As mentioned in the introduction, to predict micrometeorological data effectively, SVMs and ensemble learning have generally been used. These algorithms have higher prediction performance for micrometeorological data than traditional methods because SVMs use not only a margin maximizing algorithm whose great performance was proved by PAC learning but also the kernel trick that enables non-linear separation. On the other hand, ensemble learning provides higher generalizing capability that a single model cannot represent. In this section, a brief summary of these algorithms and some improved algorithms are given. Moreover, so that SW-SVR can draw advantages from both SVMs and ensemble learning, several problems of these algorithms for practical use are discussed.

2.1. Support vector regression

SVMs, introduced by Vapnik,(1995), have been used successfully in various fields. In the simplest case, binary classification, SVMs obtain a separating hyperplane decided by maximizing the margin. The margin means the norms between different classes. PAC learning proved that maximizing the margin produces high generalization ability. Moreover, the kernel trick enables SVMs to separate data non-linearly with low computational complexity. Various kinds of data observed in the real world are likely to have non-linear relationships. Accordingly, SVMs are used in many applications such as micrometeorological data prediction (Kisi & Cimen, 2012; Maity et al., 2010). Meanwhile, SVMs for regression, support vector regression (SVR), uses the same methodology as SVMs that have the highest generalization ability. In this section, a brief summary of SVR is given as follows.

First, the linear function for regression is given as follows:

$$f(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + b.$$

Then, as with SVMs, SVR also minimizes the norm of the weight vector $\mathbf{w}$; the L^2 norm $\|\mathbf{w}\|^2$ is often used, and minimizing $\|\mathbf{w}\|^2$ corresponds to maximizing the margin. Meanwhile, SVR tolerates prediction error ϵ. Therefore, the primal problem of SVR is shown as follows:

$$\text{minimize} \frac{\|\mathbf{w}\|^2}{2}$$
$$\text{subject to} \begin{cases} y_i - \left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b\right) \le \epsilon \\ \left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b\right) - y_i \le \epsilon. \end{cases}$$

Moreover, to take some errors into account further, the same slack variables ξ as soft margin SVMs are introduced. The slack variables mean penalties and increase in proportion to errors between true values and predicted values. The problem that the slack variables are introduced into is shown as follows:

$$\text{minimize} \frac{\|\mathbf{w}\|^2}{2} + C\sum_i \left(\xi_i + \xi_i^*\right)$$
$$\text{subject to} \begin{cases} y_i - \left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b\right) \le \epsilon + \xi_i \\ \left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b\right) - y_i \le \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0. \end{cases}$$

where the constant C means the balance between the effect of maximizing the margin and penalties. To minimize the above formula, the slack variables in the formula must also be minimized. Accordingly, the slack variables depending on the errors are shown as follows:

$$\xi_i = \begin{cases} 0 & \left(y_i - \left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b\right) \le \epsilon\right) \\ y_i - \left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b\right) - \epsilon & \text{otherwise} \end{cases}$$