Preliminary Results of Atmospheric Temperature Retrievals with Least Squares Support Vector Regression

Yao Zhigang, Chen Hongbin, Gu Hongfang

LAGEO, Institute of Atmospheric Physics, CAS, Beijing 100029, China

Abstract

The goal of this work is to investigate the performance of Least Squares Support Vector Regression (LS-SVR) for retrieving atmospheric temperatures from satellite sounding data. LS-SVR is a new regression tool and has not been widely used for geophysical parameter retrievals from satellite sounding data. Compared to artificial neural networks, LS-SVR has the advantage that it leads to a global model, which is capable of dealing efficiently with dimensional input vectors. The temperature retrievals with LS-SVR using the collocated RAOB and AMSU-A measurements over East Asia in 2002-2004 are conducted. The overall root mean square (RMS) error in the retrieved profiles of a testing dataset is remarkably smaller than the overall error using a multi-linear regression (MLR). When an offset of 0.5 K or a noise of ±0.2 K is added to all channels simultaneously, the increase in the overall RMS error is less than 0.1 K. The experiments of the variation of the training data show that for the small training dataset LS-SVR could obtain significantly more information from the sounding data than the method of the linear regression.

Introduction

One technique for retrieving temperature profiles from microwave radiances is a linear statistical inversion. But the linear statistical method is lack of the capability of retrieving temperature profiles in extreme cases and fails to address the non-linear problem. Since 1990s, Neural Networks (NNs) has been widely used and considered to be good non-linear regression methods for remote sensing. In order to retrieve atmospheric temperature profiles rapidly and accurately from microwave data, back propagation neural networks were employed in some studies (Churnside et al. 1994; Motteler et al. 1995; Butler et al. 1996; Shi 2001). However, the number of weights of the NNs is very high in cases with high dimensional input vectors. Furthermore, these weights are optimized iteratively and this procedure is repeated with different initial settings, which might lead to non-global solutions.

Recently, Support Vector Regression emerges as an alternative regression tool. SVR is a derivation of Support Vector Machines (SVM), introduced by Vapnik (1995). Least Squares Support Vector Regression (LS-SVR) is a reformation of SVR (Sukykens, 2002), which is a new regression tool and has not been widely used, especially for geophysical parameter retrievals from satellite sounding data. SVR solves regression problem by means of quadratic programming (QP). LS-SVR solves the regression problem by a set of linear equations, which is easier to use than QP. For this reason LS-SVR has a shorter computing time, as compared to SVR.

The goal of this work is to investigate the performance of Least Squares Support Vector Regression (LS-SVR) for retrieving atmospheric temperatures from satellite sounding data. Compared to artificial neural networks, LS-SVR has the advantage that it leads to a global model, which is capable of dealing efficiently with dimensional input vectors. The aim of this work is to investigate LS-SVR for performing atmospheric temperature retrievals from AMSU-A measurements.

LS-SVM regressors

Let $\{x_k, y_k\}_{k=1}^{N} \subset \Re^d \to \Re$ be the training data with inputs x_k and outputs y_k. Consider the regression model $y_k = f(x_k) + e_k$ where $x_1, \cdots x_N$ are deterministic points (fixed design), $f: \Re^d \to \Re$ is an unknown real-valued smooth function and $e_1, \cdots e_N$ are uncorrelated random errors with $E[e_i] = 0$, $E[e_i^2] = \sigma_e^2 < \infty$.

The model of a LS-SVM regressor is given as $f(x) = w^T \varphi(x) + b$ in the primal space where $\varphi(\cdot): \Re^d \to \Re^{d_f}$ denotes the potentially infinite ($d_f = \infty$) dimensional feature map. The regularized least squares cost function is given as (Sukykens, 2002)

$$\min_{w,b,e_k} J_\gamma(w,e) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{k=1}^{N} e_k^2 \qquad s.t. \quad w^T \varphi(x_k) + b + e_k = y_k, \quad k = 1, \cdots, N. \quad (1)$$

Note that the regularization constant γ appears here as in classical Tikhonov regularization. The solution corresponds with a form of ridge regression, regularization networks, Gaussian processes and Kriging, but usually considers a bias term b and formulates the problem in a primal-dual optimization context. The Lagrangian of the constrained optimization problem becomes

$$J_\gamma(w,b,e_k;\alpha_k) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{k=1}^{N} e_k^2 - \sum_{k=1}^{N} \alpha_k (w^T \varphi(x_k) + b + e_k - y_k). \qquad (2)$$

By taking the conditions for optimality $\partial J_\gamma / \partial \alpha_k = 0, \partial J_\gamma / \partial b = 0, \partial J_\gamma / \partial e_k = 0$ and $\partial J_\gamma / \partial w = 0$ and application of the kernel trick $K(x_k, x_l) = \varphi(x_k)^T \varphi(x_l)$ for all $k, l = 1 \cdots, N$ with a positive definite (Mercer) kernel K, one gets the following conditions for optimality

$$\begin{cases} y_k = w^T \varphi(x_k) + b + e_k, \quad k = 1, \cdots, N & (a) \\ \sum_{k=1}^{N} \alpha_k = 0, & (b) \\ \gamma e_k = \alpha_k, & (c) \\ w = \sum_{k=1}^{N} \alpha_k \varphi(x_k). & (d) \end{cases} \qquad (3)$$

The dual problem is summarized as follows after elimination of the variables e_k and w

$$\begin{bmatrix} 0 & 1_\upsilon^T \\ 1_\upsilon & \Omega + 1/\gamma \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \qquad (4)$$

where $\mathbf{\Omega} \in \Re^{N \times N}$ with $\Omega_{kl} = K(x_k, x_l)$, $y = [y_1, \cdots, y_N]^T$, $\mathbf{1}_\upsilon = [1, \cdots, 1]^T$ and $\alpha = [\alpha_1, \cdots, \alpha_n]^T$ for $i, j = 1; \cdots; n$. The estimated function $\hat{f}$ can be evaluated and a new point x^* by

$$\hat{f}(x^*) = \sum_{k=1}^{N} \alpha_k^* K(x_k, x^*) + b \ , \tag{5}$$

where α_k^* and b are the solution to the linear system (4).

There are several possibilities for the kernel function K. For the current experiments, LS-SVR is applied using the RBF-kernel function, because it is most widely used.

Experiments

Parameter tuning technique

In case of LS-SVR there are only two parameters to be tuned: the kernel setting (σ in case of RBF) and γ. LS-SVM model complexity (and hence its generalization performance) depends on the parameters (interaction of the kernel setting and γ). This means that separately tuning of each parameter is not feasible to find the optimal regression model. A grid search is used for tuning these two parameters. The main idea behind this method is to find the optimal parameters that minimize the prediction error of the regression model. The prediction error can be estimated by leave-n-out cross-validation on the training set. For the current study, n is set to be 1/10 of the training number. Before starting with the grid search a range for each of these parameters must be selected. The optimization ranges of these two parameters are arbitrarily defined ($\sigma = 1 - 10000$ and $\gamma = 1 - 10000$). After range selection grid search tuning is applied.

LS-SVR tuning-grid search algorithm:

1. For each set of values of the parameters, leave-n-out-cross validation on the training set is performed to predict the prediction error.
2. Select the set of values of the parameters that produced the model that gave the smallest prediction error (optimal parameter settings).
3. Train the model with the optimal parameter settings with the whole training set and test it with a test set (test is not used for training).

Software

LS-SVR is calculated using the LS-SVMlab1.5 Toolbox developed by Suykens (2003). The toolbox can be obtained from http://www.kernel-machines.org. The Toolbox works under Matlab (The Mathworks, Inc.).

Performance

The Root-Mean-Square-Error (RMSE) is used as performance criterion in cross-validation and also for predicting the test set. The RMSE of retrievals is defined as

$$RMSE = \sqrt{\frac{1}{N_s} \sum_{i=1}^{N_s} (X_{RAOB} - X_{RETR})^2} \ , \tag{6}$$

where X_{RAOB} and X_{RETR} are the radiosonde observation and retrieved temperatures respectively, and N_s is the total number of comparisons.

Data

The data used for the current study are from the direct broadcast data of NOAA-16 received at Beijing, China. The maximum coverage of the data extends from 15ºN to 60º N and 80º E to 140º E. AMSU-A has 20 microwave channels, the weighting functions of which are from the surface to the 0.1-hPa level. Due to the fact that the radiosonde data generally report the parameters below 100-hPa, only AMSU-A channels 1-10 and 15 measuring the radiances mainly emitted by the atmosphere below the 100-hPa level are used in this study. Consequently, the input vector comprises twelve elements, which are the brightness temperatures at the eleven selected channels of AMSU-A and the secant of the local satellite zenith angle of observation

The AMSU-A 1d data are matched with the radiosonde observations for the period of Jan. 2002 to Sep. 2004 over the land. The criteria for selecting AMSU-A measurements with collocated radiosonde data are based on the following: 1) The absolute distance between the position (latitude and longitude) of the radiosonde and the ATOVS retrieval FOV (Field of View) is less than 0.5º (the center of FOV is chosen to represent the position of the ATOVS retrieval FOV). 2) The time difference between radiosonde and ATOVS measurements is less than 1.5 h. 3) The radiosonde observations from reporting stations with terrain heights less than 500 m are selected. 4) Brightness-temperature records of the AMSU-A are complete. 5) The radiosonde observation has no missing layers from 1000- to 100-hPa pressure levels. Based on these criteria, there are 7593 collocated samples available over the land, including 4649 from Jan. 2002 to Dec. 2003 and 2944 from Jan. 2004 to Sep. 2004. The former dataset (dataset A) is used to train the regression model, and the latter (dataset B) is used to test the model. As a consequence, the data used in the training set are not included in the testing set. It is very important to evaluate the capability of the retrieval methods using data independent of the training data.

Results and discussion

For saving computing time, only 400 collocated pairs for training are uniformly taken from the dataset A. After the grid search, the dataset B is used to test the obtained optimal regression model. Furthermore, in order to examine the difference of a LS-SVM approach from a linear regression approach, the temperature retrieval using the multi-linear regression (MLR) is carried out. The overall RMS retrieval error of LS-SVR is 2.04 K, and the overall error of MLR is 2.41 K. The RMS retrieval errors at different pressure levels are shown in Fig. 1. From Fig.1, it is clear that LS-SVR could obtain better results than the multi-linear regression at all levels, especially at the 200-, 250-, 300-, and 850-hPa levels.

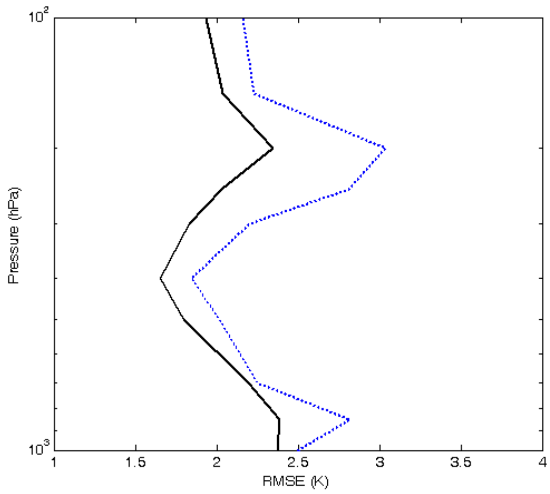

Fig. 1 Comparison of the RMS retrieval errors of LS-SVM with MLR (Dotted: MLR; Solid: LS-SVR)

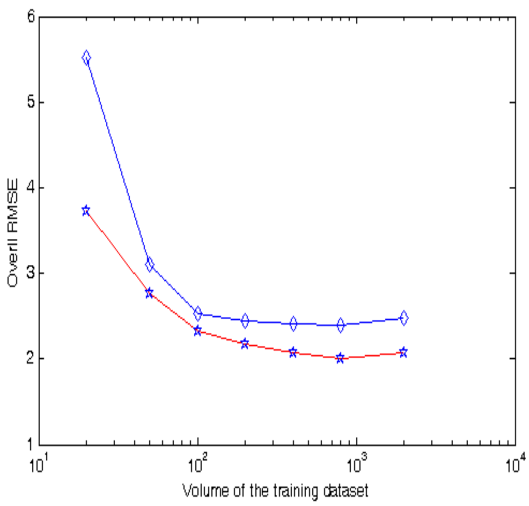

Fig.2 Comparison of the overall RMS errors of LS-SVM with MLR for the different volume of the training dataset (Blue: MLR; Red: LS-SVR)

Additionally, in order to evaluate the effects of the volume of training dataset on the retrieval, seven experiments are conducted separately. The number of the collocated pairs for training is 20, 50, 100, 200, 400, 800, and 2000, respectively. The dataset B is used to test the obtained parameters. Similarly, the multi-linear regression method is also applied. The overall RMS retrieval errors are shown in Fig. 2. The comparison shows that for the small training dataset LS-SVR could obtain significantly more information from the sounding data than the method of the linear regression.

Effects of noise and calibration offset

Because there may be some difficulties during the long operation period, it is interesting and timely to investigate the effect on retrieval accuracy of some failure modes of AMSU-A. Two types of experiments are conducted to examine the retrieval accuracy of the regression model trained using all channels under two possible system problems: equal offset of all channels and equal noise in all channels. We use the same model trained from only 400 collocated pairs for following retrieval experiments.

In the first series of experiments, the effect of simultaneously offsetting all brightness temperatures by a constant is determined. For increasing offsets of 0.2, 0.5, 1.0, 1.5, and 2.0 K, the increase in overall retrieval error is shown in Table 1. The results show that the LS-SVR method is relatively immune to offset when the offset is less than 0.5 K. From Table 1, one can also see that the overall retrieval error of LS-SVR model with the offset of 2.0 K is slightly less than that of MLR model, although the increase of the error of the former is larger than that of the latter.

Table 1. Variation of the overall error with the offset
(Training: 400; Testing: 2944)

	0.0	0.2	0.5	1.0	1.5	2.0
MLR	2.41	+0.02	+0.07	+0.21	+0.41	+0.66
LS-SVR	2.07	+0.02	+0.08	+0.29	+0.59	+0.95

In the noise experiments, the retrieval accuracies are examined while all channels become noisy. We follow the same procedure as in the offset studies. The effect of adding these amounts of random noise to all channels simultaneously is considered. In this case, six testing files are made in which all channels suffer additional uniform random noise centered about zero with maximum values in the ranges (-0.1, 0.1), (-0.2, 0.2), (-0.3, 0.3), (-0.5, 0.5), (-0.8, 0.8) or (-1.0, 1.0) K. The increase in overall retrieval error compared to the error without the noise is shown in Table 2. The results indicate that the effect is neglectable when the additional noise is less than 0.2K. The results also show that the stability of the LS-SVR model trained is comparable with that of the MLR model.

Table 2. Variation of the overall error with the additional noise
(Training: 400; Testing: 2944)

	0.0	±0.1	±0.2	±0.3	±0.5	±0.8	±1.0
MLR	2.41	+0.01	+0.05	+0.10	+0.25	+0.60	+0.88
LS-SVR	2.07	+0.01	+0.05	+0.11	+0.25	+0.61	+0.89

Conclusions and future work

The goal of this work is to investigate the feasibility of LS-SVR for retrieving atmospheric temperatures from satellite radiances.

In general, it can be concluded that LS-SVR will give better results than MLR for temperature retrievals. The preliminary experimental results show that LS-SVM should be an effective tool to model the complicated relationship between geophysical parameters and satellite remote sensing data.

We plan to apply LS-SVR to retrieve temperature and other geophysical parameters simultaneously from the infrared and microwave data.

Acknowledgement

We thank the National Satellite Meteorological Center, China Meteorological Administration for providing ATOVS/L1d data. This work was supported by the China National Key Basic Research Program No. 2001CB309402.

Reference

Churnside, J. H., T. A. Stermitz, and J. A. Schroeder, 1994: Temperature profiling with neural network inversion of microwave radiometer data. *J. Atmos. Ocean Tech.*, **11**, 105-109.

Motteler, H.E., L.L. Strow, and L. Mcmillin et al., 1995: A comparison of neural networks and regression-based methods for temperature retrievals. *Appl. Opt.*, **34**, 5390-5397.

Butler, C. T., R. v. Z. Meredith, and A. P. Stogryn 1996: Retrieving atmospheric temperature parameters from DMSP SSM/T-1 data with a neural network. *J. Geophys. Res* , **101**, 7075-7083.

Shi Lei, 2001: Retrieval of atmospheric temperature profiles from AMSU-A measurement using a neural network approach. *J. Atmos. Ocean. Tech.*, **18**, 340-347.

Vapnik, V. 1995, *The Nature of Statistical Learning Theory*. New York: Springer, USA.

Sukykens, J.A.K., T. van Gestel, J. De Brabanter, B. De Moor and J. Vandewalle, 2002. *Least Squares Support Vector Machines*. London, UK: World Scientific Publishing Co. Pte. Ltd.

Sukykens, J.A.K., K. Pelckmans, T. van Gestel, J. De Brabanter, L. Lukas, B. Hamers, B De Moor, J. Vandewalle, *LS-SVM Toolbox*, version 1.5, http://www.esat.kuleuven.ac.be/sista/lssvmlab/, 2003: Leuven, Belgium.