Fast Optimization of Hyperparameters for Support Vector Regression Models with Highly Predictive Ability

Hiromasa Kaneko and Kimito Funatsu

Department of Chemical System Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

Corresponding author: Kimito Funatsu

Tel: +81-3-5841-7751

Fax: +81-3-5841-7771

E-mail: funatsu@chemsys.t.u-tokyo.ac.jp

Abstract

Support vector regression (SVR) attracts much attention in chemometrics as a nonlinear regression method due to its theoretical background. In SVR modeling, three hyperparameters must be set beforehand. The optimization method based on grid search (GS) and cross-validation (CV) is employed normally in the selection of the SVR hyperparameters. However, this takes enormous time. Although theoretical techniques exist to decide the values of the SVR hyperparameters, predictive ability of SVR models is not considered in the decision. We therefore proposed a method based on the GS and CV method and theoretical techniques for fast optimization of the SVR hyperparameters, considering predictive ability of SVR models. After values of two hyperparameters are decided theoretically, each hyperparameter is optimized independently with GS and CV. The highly predictive ability of SVR models and small computational time for the proposed method are confirmed through the case studies using real data sets.

Keywords: Support vector regression; Hyperparameter; Optimization; Predictive ability; Computational time.

1 Introduction

Chemometrics aims to solve chemistry problems using statistical methods. There is a great deal of previous research in this field, revolving around topics such as quantitative structure-activity relationships (QSARs) and quantitative structure-property relationships (QSPRs), process and product designs and soft sensors.

In this paper, we focus on regression analysis. Regression models are constructed between explanatory variables X and an objective variable y by using regression methods. When relationships between X and y are linear, linear regression methods can be used to construct regression models. There exist many types of linear regression methods such as ordinary least squares, ridge regression [1], least absolute shrinkage and selection operator [2], principal component regression [3] and partial least squares (PLS) [4].

However, when relationships between X and y are nonlinear, linear regression models cannot represent these relationships. Nonlinear regression methods such as back-propagation neural network [5], counter-propagation neural network [6], kernel PLS [7], Gaussian process [8] and support vector regression (SVR) [9] are required. In this study, we focus on SVR because of its theoretical background and the Gaussian kernel is usually used in SVR modeling.

Although nonlinear regression models can be constructed by using the SVR method, over-fitting, i.e., [the propensity of the model to reproduce better the training data to the detriment of its generalization](#), is a crucial problem and should be cared in model construction. Figure 1 shows the basis concept of over-fitting. The regression model in Figure 1(b) is fitting to the training data very much, but the prediction errors for new data are large. [The simpler, more regular model illustrated in Figure 1\(a\) is a better choice. It is less accurate in reproducing the training data but performs much better in producing useful estimates for the test data.](#) Although there are nonlinear regression models in Figure 1, multi-linear regression methods such as a PLS and an SVR using a linear kernel have possibility of over-fitting as well.

To prevent over-fitting, values of three hyperparameters in the SVR model, the details of which are shown in Appendix A, must be selected carefully. After the selection of the hyperparameter-values, predictive ability of the SVR model is checked. The comprehensive grid search (GS) is employed to investigate all the combinations of the hyperparameters and cross-validation (CV) is used to check predictive ability of SVR models. The hyperparameter combination having the maximum r^2 -value with CV, which is the r^2_{CV} -value in Appendix B, is selected as the final hyperparameter-values.

To prevent overestimating predictive ability of SVR models, N -fold CV is required [10,11].

By using the GS and CV method, the appropriate hyperparameter-values can be selected while considering predictive ability of SVR models. However, this procedure takes enormous computational time although recently the root-mean-square error for the midpoints between k -nearest-neighbor data points [12] has been proposed to reduce computational time in model validation and [Bergstra and Bengio proposed random search for hyperparameter optimization as a more efficient method than GS \[13\]](#). Table 1 shows the candidates of three SVR hyperparameters C , ε and γ . When C has 26 candidates; ε has 16 candidates; and γ has 31 candidates, N -fold CV has to be performed 12,896 ($=26 \times 16 \times 31$) times. In addition, for single N -fold CV, SVR models must be constructed N times. When we handle big data [14,15], i.e., very large and complex data sets, we cannot construct SVR models so many times.

On the other hand, the SVR hyperparameter-values can be decided theoretically. C is given based on the variance in a y -variable and ε is given using the calculation results with k -nearest neighbor algorithm [16]. γ is given by maximizing the variance in a dataset using the Gaussian kernel function [17]. This procedure takes little time since no SVR models are constructed in the hyperparameter decision. However, predictive ability of SVR models is not considered in the selection of the hyperparameter-values.

In GS and CV, it takes large time although predictive ability of SVR models can be considered. In theoretical decision (ThD), predictive ability of SVR models is not considered while it takes little time. We therefore propose a method for fast optimization of the SVR hyperparameters considering the predictive ability by combining the GS and CV method and the ThD method. After the SVR hyperparameters are decided theoretically, the optimization using GS and CV is performed for each hyperparameter in order. The SVR model with the hyperparameter-values optimized by the proposed method has the same predictive ability as that of the GS and CV method, and the computational time can be drastically reduced compared with that of the GS and CV method because the theoretical techniques are employed in the hyperparameter selection.

To investigate predictive ability of SVR models whose hyperparameter-values are selected with the proposed method and to check the computational time to decide the final hyperparameter-values, we analyze a real QSPR dataset and a real QSAR dataset. The GS and CV method and the ThD method are compared in terms of predictive ability of SVR models and computational time.

2 Method

In SVR modeling, the hyperparameters C , ε and γ must be set beforehand. In this paper, after the GS and CV method and the ThD method, we explain the proposed method combining the GS and CV method and the ThD method.

2.1 GS and CV

The comprehensive combinations of the hyperparameters in Table 1 are checked in terms of predictive ability of the SVR models. CV on training data is a standard method to investigate the predictive ability. CV is a technique evaluating the predictive ability of statistical models by using only the training data. N -fold CV is recommended rather than leave-one-out CV to prevent over-fitting. In N -fold CV, the training data are randomly divided into N groups, in which the numbers of data are the same as far as possible. One group is then used as the data to validate the model constructed using the data from the other ($N-1$) groups. This procedure is repeated N times so that the data of each of the N groups are used once as the validation data. Finally predicted rather than calculated values of $\mathbf{y}$ can be obtained. In this paper, N is set as 5, which is a standard value.

The predictive ability of the SVR model can be considered in the hyperparameter setting process. However, this procedure is time-consuming. In Table 1, C has 26 candidates, ε has 16 candidates, and γ has 31 candidates, and then, CV has to be performed 12,896 ($=26 \times 16 \times 31$) times. Additionally, for single N -fold CV, SVR models must be constructed N times.

2.2 ThD

Cherkassky and Ma proposed the theoretical decision method to decide the hyperparameters C and ε of a SVR model [16]. The values of C and ε are calculated as follows:

$$C = \max\left(\left|m_y + 3s_y\right|, \left|m_y - 3s_y\right|\right), \quad (1)$$

$$\varepsilon = 3\sigma \sqrt{\frac{\ln n}{n}}. \quad (2)$$

Here m_y is the average of $\mathbf{y}$ and s_y is the standard deviation of $\mathbf{y}$. σ in Eq. (2) is estimated with the k -nearest-neighbor's method as follows:

$$\sigma^2 = \frac{n^{1/5}k}{n^{1/5}k-1} \cdot \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (3)$$

Here y_i with the marker above it is an estimated y -value and n is the number of training data. k is recommended to be 3 [16].

Tang *et al.* proposed to optimize the hyperparameter γ in the Gaussian kernel function [17] (see Appendix A). γ is given by maximizing variance in a dataset using the Gaussian kernel function. The variance σ_γ^2 is calculated as follows:

$$\sigma_\gamma^2 = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \left(K(\mathbf{x}_i, \mathbf{x}_j) - m_\gamma \right)^2, \quad (4)$$

where

$$m_\gamma = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n K(\mathbf{x}_i, \mathbf{x}_j). \quad (5)$$

K is the Gaussian kernel function in eq. (A.4) in Appendix A and $\mathbf{x}_i$ and $\mathbf{x}_j$ are training data. In the reference [17], the golden section search is used to find the optimal γ .

Although this theoretical decision takes little time because CV is not required and no SVR model is constructed, predictive ability of SVR models is not considered through the decision procedure.

2.3 Proposed method

Figure 2 shows the flow of the proposed method. First, C and γ are decided theoretically because γ can be decided irrespective of SVR modeling and decision of C takes less time than that of ε . Then, using decided C and γ , only ε is optimized using GS and CV. Next, using decided γ and optimized ε , only C is optimized with GS and CV. Finally, using optimized C and ε , only γ is optimized with GS and CV.

For the previous GS and CV method, there were 12,896 (=26×16×31) candidates in Table 1, but the proposed method needs only 73 (=26+16+31) candidates because C , ε and γ can be optimized using GS and CV independently. This is only 0.6% of the number of hyper-parameters values to test compared to standard grid search optimization. The proposed method can optimize three SVR hyperparameters with little time while considering predictive ability of SVR models.

Bergstra and Bengio proposed random search for hyperparameter optimization

and showed that random search is more efficient than GS [13]. In each optimization step in Figure 2, random search can be replaced with GS.

When the number of data is large in a database, external validation, in which a fraction of a database is used as an external validation data set, is employed to validate predictability of regression models instead of CV. The proposed method can be easily combined with external validation. By replacing CV with external validation in Figure 2, the SVR hyperparameters are optimized to have high predictability for an external data set. Midpoints between k-nearest-neighbor data points [12] can be also used in validation.

3 Results and Discussion

The QSPR data of aqueous solubility and the QSAR data of pIGC₅₀ were used to optimize the SVR hyperparameters and to validate the SVR models constructed with the optimized hyperparameters-values. GS and CV, ThD and the proposed method were compared as the optimization methods and the predictive ability of the SVR models and the computational time in optimization were discussed. The candidates of C , ε and γ are shown in Table 1.

The configurations of the computer are given as follows:

OS: Windows 7 Professional (64 bit)

CPU: Intel(R) Xeon(R) X5690 3.47GHz; RAM: 48.0 GB

Software: MATLAB R2014a.

3.1 QSPR study on aqueous solubility

We analyzed the aqueous solubility data investigated by Hou [18]. Aqueous solubility is expressed as logS, where S is the solubility (in moles per liter) at a temperature of 20–25°C. The data set includes 1,290 diverse compounds.

2,232 molecular descriptors were calculated on this data set using the Dragon 6.0 software [19]. First, descriptors with the same values for 100% of the molecules were removed. Then, when the correlation coefficient of a pair of descriptors was greater than 0.9, one of the descriptors was removed, which left 225 descriptors.

The number of training data is 500 and that of test data is 790. The training data were selected randomly 10 times and the SVR hyperparameters were optimized each time. Accordingly, the test data were also changed 10 times.

Figure 3 shows the r^2_p -values for test data sets (see Appendix B for r^2_p) when the SVR hyperparameters are optimized using GS and CV, ThD and the proposed method. The higher r^2_p -values mean the higher predictive ability of SVR models. In

Figure 1, there are 10 r_p^2 -values for each method because training data are changed 10 times. When SVR parameters were decided theoretically, the r_p^2 -values were low and accordingly the predictive ability of the SVR models was low, which was because the predictive ability of SVR models was not considered in the decision of the SVR hyperparameters in ThD. On the other hands, when SVR parameters are optimized with the proposed method, the r_p^2 values were high, which was almost the same level as that in GS and CV. This indicates that the predictive ability of the SVR models was high for both the proposed method and the GS and CV method.

The plots of observed and predicted logS for a test data set are shown in Figure 4. For ThD, [the errors are large resulting in many data points lying far from the diagonal](#). However, for the GS and CV method and the proposed method, the data are close to the diagonals in Figure 4(a)(c), reflecting the high prediction ability of the SVR models. In addition, when values of logS are large, the prediction errors for the proposed method are lower than those for the GS and CV method. We confirmed that the proposed method could optimize the SVR hyperparameters whose SVR models had the highly predictive ability.

Table 2 shows computational time for each method. [There are no outliers of computational time for each method](#). The GS and CV method has the largest time as expected. When this computational time is 100 %, the computational time of ThD is 0.04% and that of the proposed method is 0.2%. By using the proposed method, the computational time could be reduced dramatically, compared to that of the GS and CV method. The decrease ratio was less than the theoretical one (0.6%). The computational time in SVR modeling depends on values of the SVR hyperparameters and inappropriate hyperparameter-values need large computational time. By theoretically deciding appropriate hyperparameters first, overall computational time in SVR modeling could be reduced largely.

3.2 QSAR study on pIGC₅₀

We analyzed data downloaded from the Environmental Toxicity Prediction Challenge 2009 website [20]. This is an online challenge that invites researchers to predict the toxicity of molecules against *T. Pyriformis*, expressed as the logarithm of 50% growth inhibitory concentration (pIGC₅₀). The data set consists of 1,093 compounds.

2,232 molecular descriptors were calculated on this data set using the Dragon 6.0 software [19]. Descriptors with the same values for 100% of the molecules were removed. When the correlation coefficient of a pair of descriptors was greater than 0.9,

one of the descriptors was removed, which left 333 descriptors.

The training data set comprised 500 molecules and the remaining 593 molecules were the test data set. We randomly selected training data 10 times and accordingly test data also changed 10 times.

r^2_p -values for each method is shown in Figure 5. As in the case of the QSPR data analysis, when the SVR hyperparameter-values were decided theoretically, the prediction accuracy was very low. However, when the hyperparameters were selected with the GS and CV method or the proposed method, the r^2_p -values were high and the prediction accuracy was also high. Figure 6 shows the plots of observed and predicted pIGC₅₀ for each method. For ThD, the errors are large resulting in many data points lying far from the diagonal. On the other hands, for the GS and CV method and the proposed method, the data are close to the diagonals and overall prediction errors are small, which means high predictive ability of the SVR models. The appropriate SVR hyperparameters with which we could construct the predictive SVR model could be selected by using the proposed method.

Table 3 shows computational time for each method. There are no outliers of computational time for each method. The GS and CV method has the largest time as well as it does in the QSPR data analysis. When this computational time is 100 %, that of the ThD method is 0.07% and that of the proposed method is 0.21%. By using the proposed method, the computational time could be reduced dramatically. As in the QSPR analysis, the actual reduction ratio was less than the theoretical one. This is because the computational time for SVR modeling depends on the SVR hyperparameters and the proposed method could prevent the long computational time in SVR modeling by deciding appropriate hyperparameters with theoretical techniques beforehand.

We confirmed that the proposed method could select the appropriate SVR parameters quickly and could construct the SVR model having the highly predictive ability.

4 Conclusions

In this paper, a fast optimization method of the SVR hyperparameters is proposed by combining the theoretical techniques and the GS and CV method. After C and γ are decided theoretically, ε , C and γ are optimized with GS and CV independently. Therefore the proposed method can consider predictive ability of SVR models and needs little time for optimizing the SVR hyperparameters. Through the case studies using the QSPR data and QSAR data, it was confirmed that the SVR hyperparameters

with which the predictive SVR models could be constructed were able to be quickly selected by using the proposed method. [As is mentioned in “2.3 Proposed method”, the proposed method can be combined with external validation instead of CV.](#)

Because the SVR hyperparameters can be optimized with little time using the proposed method, nonlinear SVR models can be constructed many times within a reasonable time. Therefore SVR models can be easily applied to ensemble learning, variable selection, sample selection and so on.

The proposed optimization method of hyperparameters can be applied to classification and regression method based on Gaussian kernel such as support vector machine [21], kernel PLS [7] and Gaussian process [8].

Acknowledgements

The authors acknowledge the support of the Core Research for Evolutionary Science and Technology (CREST) project “Development of a knowledge-generating platform driven by big data in drug discovery through production processes” of the Japan Science and Technology Agency (JST). H. Kaneko is grateful for the financial support of the Japan Society for the Promotion of Science (JSPS) through a Grant-in-Aid for Young Scientists (B) (No. 24760629).

References

- [1] I.E. Frank, J.H. Friedman, A statistical view of some chemometrics regression tools, *Technometrics* 35 (1993) 109–135.
- [2] R. Tibshirani, Regression shrinkage and selection via the lasso. *J. R. Stat. Soc.* 58 (1996) 267–288.
- [3] S. Wold, Principal component analysis. *Chemom. Intell. Lab. Syst.* 2 (1987) 37–52.
- [4] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* 58 (2001) 109–130.
- [5] J. Rantanen, E. Rasanen, O. Antikainen, J.P. Mannermaa, J. Yliruusi, In-line moisture measurement during granulation with a four-wavelength near-infrared sensor: an evaluation of process-related variables and a development of non-linear calibration model, *Chemom. Intell. Lab. Syst.* 56 (2001) 51–58.
- [6] M. Arakawa, K. Hasegawa, K. Funatsu, QSAR study of anti-HIV HEPT analogues based on multi-objective genetic programming and counter-propagation neural network, *Chemom. Intell. Lab. Syst.* 83 (2006) 91–98.
- [7] K. Kim, J.M. Lee, I.B. Lee, A novel multivariate regression approach based on kernel partial least squares with orthogonal signal correction, *Chemom. Intell. Lab.*

Syst. 79 (2005) 22–30.

- [8] T. Chen, J. Morris, E. Martin, Gaussian process regression for multivariate spectroscopic calibration, *Chemom. Intell. Lab. Syst.* 87 (2007) 59–67.
- [9] H.D. Li, Y.Z. Liang, Q.S. Xu, Support vector machines and its applications in chemistry, *Chemom. Intell. Lab. Syst.* 95 (2009) 188–198.
- [10] K. Baumann, Cross-validation as the objective function for variable-selection techniques, *Trac-Trend Anal. Chem.* 22 (2003) 395–406.
- [11] A. Golbraikh, A. Tropsha, Beware of q^2 !, *J. Mol. Graph Model.* 20 (2002) 269–276.
- [12] H. Kaneko, K. Funatsu, Criterion for evaluating the predictive ability of nonlinear regression models without cross-validation, *J. Chem. Inf. Model.* 53 (2013) 2341–2348.
- [13] J. Bergstra, Y. Bengio, Random search for hyper-parameter optimization, *J. Mach. Learn. Res.* 13 (2012) 281–305.
- [14] D. Howe, M. Costanzo, P. Fey, T. Gojobori, L. Hannick, W. Hide, D.P. Hill, R. Kania, M. Schaeffer, S. St Pierre, S. Twigger, O. White, S.Y. Rhee, Big data: The future of biocuration, *Nature.* 455 (2008) 47–50.
- [15] J. Camacho, Visualizing big data with compressed score plots: Approach and research challenges, *Chemom. Intell. Lab. Syst.* 135 (2014) 110–125.
- [16] V. Cherkassky, Y. Ma, Practical selection of SVM parameters and noise estimation for SVM regression, *Neural Networks* 17 (2004) 113–126.
- [17] Y. Tang, W. Guo, J. Gao, Efficient model selection for support vector machine with Gaussian kernel function, *Proceedings of Computational Intelligence and Data Mining* (2009) 40–45.
- [18] T.J. Hou, K. Xia, W. Zhang, X.J. Xu, ADME evaluation in drug discovery. 4. prediction of aqueous solubility based on atom contribution approach. *J. Chem. Inf. Comput. Sci.* 44 (2004) 266–275.
- [19] http://www.taletе.mi.it/products/dragon_description.htm
- [20] <http://www.cadaster.eu/node/65>
- [21] C.M. Bishop, *Pattern recognition and machine learning*, Springer, New York, 2006.
- [22] <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

Appendix A: SVR

The SVR method applies an SVM to regression analysis and can be used to construct non-linear models by applying a kernel trick along with the SVM. The primal form of the SVR can be shown to be the following optimization problem.

Minimize

$$\frac{1}{2}\|\mathbf{w}\|^2 + C \sum_{i=1}^N |y_i - f(\mathbf{x}_i)|_{\varepsilon}, \quad (\text{A.1})$$

where y_i and $\mathbf{x}_i$ are training data, f is an SVR model, $\mathbf{w}$ is a weight vector, ε is a threshold, N is the number of training data, and C is a penalizing factor that controls the trade-off between the model complexity and the training errors. The second term of Eq. (A.1) is the ε -insensitive loss function and is given as follows:

$$|y_i - f(\mathbf{x}_i)|_{\varepsilon} = \max(0, |y_i - f(\mathbf{x}_i)| - \varepsilon). \quad (\text{A.2})$$

By minimization of Eq. (A.1), we can construct a regression model with a good balance between its generalization capabilities and its ability to adapt to the training data. A y -value predicted by inputting data $\mathbf{x}$ is represented as follows:

$$f(\mathbf{x}) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b, \quad (\text{A.3})$$

where b is a constant and K is a kernel function. The kernel function in our application is a radial basis function:

$$K(\mathbf{x}_i, \mathbf{x}) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}\|^2), \quad (\text{A.4})$$

where γ is a tuning parameter that controls the width of the kernel function. From eqs. (A.1) and (A.2), α_i and α_i^* in eq. (A.3) can be obtained by minimizing the equation given as:

$$\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N K_{ij} (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) - \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) + \varepsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*), \quad (\text{A.5})$$

subject to

$$0 \leq \alpha_i, \alpha_i^* \leq C \quad i = 1, 2, \dots, N, \quad (\text{A.6})$$

$$\sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0. \quad (\text{A.7})$$

K_{ij} in Eq. (A.5) is represented as follows:

$$K_{ij} = K(\mathbf{x}_i, \mathbf{x}_j). \quad (\text{A.8})$$

LIBSVM [22] was used as the machine learning software.

Appendix B: Statistics

To construct a highly predictive model, hyperparameter-values must be chosen appropriately. The r_{CV}^2 values are used as measured and defined as follows:

$$r_{\text{CV}}^2 = 1 - \frac{\sum_{i=1}^n (y_{\text{obs}}^{(i)} - y_{\text{pred}}^{(i)})^2}{\sum_{i=1}^n (y_{\text{obs}}^{(i)} - \bar{y}_{\text{obs}})^2} \quad (\text{B.1})$$

Where y_{obs} is the measured y -value and y_{pred} is the predicted y -value in the cross-validation procedure. In this study, the 5-fold cross validation method is used in the calculation of y_{pred} . r_{CV}^2 represents the predictive accuracy of the constructed models. Values close to one for r_{CV}^2 are favorable. The r_{CV}^2 -value depends on the number of folds used in cross-validation; therefore, performances of regression models can be compared only if they are measured using the same cross-validation protocol. The Pearson's coefficient can be determined on a test data set in that case it is noted r_p^2 :

$$r_p^2 = 1 - \frac{\sum_{i=1}^{n_{\text{test}}} \left(y_{\text{obs,test}}^{(i)} - y_{\text{pred,test}}^{(i)} \right)^2}{\sum_{i=1}^{n_{\text{test}}} \left(y_{\text{obs,test}}^{(i)} - \bar{y}_{\text{obs,test}} \right)^2}, \quad (\text{B.2})$$

where n_{test} is the number of data in a test data set and $y_{\text{obs,test}}$ with the marker above it is the average value of the y -values on a test data set.

Figures captions

Figure 1. Basic concept of an appropriate regression model and an over-fitted regression model.

Figure 2. Flow of the proposed method.

Figure 3. r_p^2 -values with the QSPR data.

Figure 4. Observed and predicted logS for a test data set in SVR modeling when the hyperparameters are optimized with each method with the QSPR data.

Figure 5. r_p^2 -values with the QSAR data.

Figure 6. Observed and predicted pIGC₅₀ for a test data set in SVR modeling when the hyperparameters are optimized with each method with the QSAR data.

Table 1. Candidates of C , ε and γ in SVR modeling.

C	$2^{-10}, 2^{-9}, \dots, 2^{14}, 2^{15}$	26 candidates
ε	$2^{-15}, 2^{-14}, \dots, 2^{-1}, 2^0$	16 candidates
γ	$2^{-20}, 2^{-19}, \dots, 2^9, 2^{10}$	31 candidates

Table 2. Computational time in the hyperparameter optimization with the QSPR data.

	GS and CV	ThD	Proposed method
Average [s]	6801 (100 %)	3.0 (0.04%)	13.3 (0.20%)
Standard deviation [s]	53	0.75	0.77

Table 3. Computational time in the hyperparameter optimization with the QSAR data.

	GS and CV	ThD	Proposed method
Average [s]	8400 (100 %)	6.1 (0.07%)	17.9 (0.21%)
Standard deviation [s]	78	0.1	0.1

Figure(s)

Figure 1. Basic concept of an appropriate regression model and an over-fitted regression model.

Figure 2. Flow of the proposed method.

Figure 3. r_p^2 -values with the QSPR data.

Figure 4. Observed and predicted logS for a test data set in SVR modeling when the hyperparameters are optimized with each method with the QSPR data.

Figure 5. r_p^2 -values with the QSAR data.

Figure 6. Observed and predicted $pIGC_{50}$ for a test data set in SVR modeling when the hyperparameters are optimized with each method with the QSAR data.