Transactions of the Korean Nuclear Society Autumn Meeting
Gyeongju, Korea, October 26-27, 2017

Support Vector Machines for Event Classification and Regression Analysis

Young Do Koo [a] and Man Gyun Na [a*]
[a]*Department of Nuclear Engineering, Chosun Univ., 309 Pilmun-daero, Dong-gu, Gwangju, Korea 61452*
[*]*Corresponding author: magyna@chosun.ac.kr*

1. Introduction

The algorithm that gives computers the learning capability, and accordingly enable data-driven decision, estimation, and clustering is based on machine learning. The purpose of machine learning is to develop the algorithm that performance P on task T improves with experience E. In this sense, the studies for developing the optimal machine learning algorithm such as artificial neural networks (ANNs), Bayesian inference, fuzzy inference, and support vector machines (SVMs) have been carried out.

Among them, especially, the studies on application of SVMs to classification and regression problems is described in this paper. The current embodiment of SVMs was proposed by C. Cortes and V. Vapnik in 1995 [1] and it is an algorithm with a neural network structure based on statistical learning theory. These SVMs have been generally used for event classification and identification. In addition, after the introduction of Vapnik's ε-insensitive loss function [2,3], SVMs have been widened to be used to solve the nonlinear regression problems. In other words, SVMs are supervised learning models related to the learning algorithm that analyzed the data used for classification and regression problems. Additionally, using the kernel function [1], SVMs can effectively perform the nonlinear classification and regression analysis.

There are studies using these SVMs in instrumentation and control field of nuclear power plants (NPPs). First of all, 7 transients of NPPs were classified and identified [4]. Furthermore, golden time for accident recovery [5], power peaking factor (PPF) [6], departure from nuclear boiling ratio (DNBR) [7], residual stress of welding metal [8], and loss of coolant accident (LOCA) break size [9] were estimated using SVM models.

2. Support Vector Machines

As an artificial intelligence method applied in nuclear industrial fields, SVMs are learning tools that use a hypothesis space of a linear function in a higher dimensional space, which are trained through a learning algorithm originated from statistical learning theory. Although the structure of SVMs and artificial neural networks (ANNs) are similar, they are differ with the aspect of learning method or risk minimization optimization [10]. SVMs utilizes a structural risk minimization (SRM) principle to make a minimum of the upper bound on the expected risk [11]. This difference of the risk minimization enable SVMs to have better generalization performance than ANNs [11]. A SRM principle is depicted in Fig. 1.

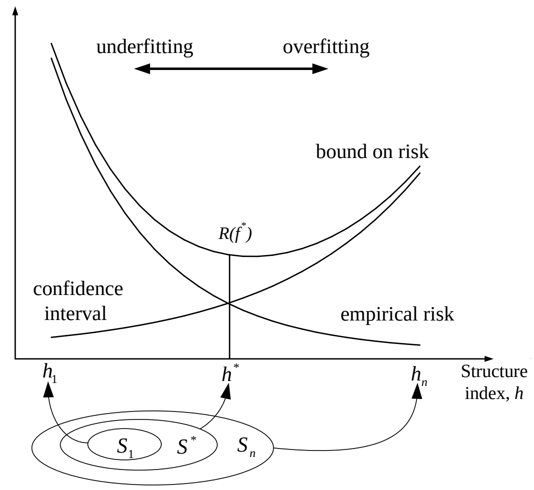

Fig. 1. An illustration of the SRM principle [11].

The SVM models for classification and regression analysis are established using N learning data. The learning data are indicated as $T=\{(\mathbf{x}_i,\, y_i)\}_{i=1}^{N}$ where $\mathbf{x}_i \in R^m$ means the sample data vector and y_i means a class $y_i \in \{+1,-1\}$, from which a relationship between an input and output is learned.

In addition, in previous studies, the slack variable δ associated with the learning efficiency of the SVM model was used. Generally, it is regarded that all the data are not be able to be precisely separated. Thus, the slack variable contributes to establishing the optimal SVM models by making a data selection area wider.

Nonlinear classification and regression problems can be changed into a linear analysis using the kernel function. That is, this is to nonlinearly map the data from the initial space into a kernel-induced higher dimensional space. Because of the best performance in aforementioned studies, the radial basis function was generally used as follows:

$$K(\mathbf{x}_i,\mathbf{x}) = \exp\left(-\frac{(\mathbf{x}-\mathbf{x}_i)^T(\mathbf{x}-\mathbf{x}_i)}{2\sigma^2}\right) \quad (1)$$

2.1 SVMs for Event Classification

In case that SVMs are applied to a classification problem, it is called support vector classification (SVC). Classification problems using SVMs can be commonly considered as a two-class classification problem as shown in Fig. 2. Two-class classification method can establish the decision boundary to separate the data vectors into one of two classes based on a learning data set of which classification is known as a *priori*. A decision boundary dividing the classes is usually expressed as follows:

$$\mathbf{w}\cdot\mathbf{x}+b=0 \tag{2}$$

where vector **w** and bias *b* determine the decision boundary.

Even though, however, several classifiers that separate the data into two-class exist, only the one classifier can become the optimal separating hyperplane, which has a the widest margin that is the interval between the classifier and the nearest data points of each class. The optimal separating hyperplane can be established by minimizing the following function:

$$\Phi(\mathbf{w})=\frac{1}{2}\left|\mathbf{w}\right|^2 \tag{3}$$

As shown in Fig. 2, magnitude of **w** has to be minimized to make the margin maximum. Finally, the optimal separating hyperplane of SVC function using the slack variables and kernel function becomes:

$$f(x)=\mathrm{sgn}\left(\sum_{i\in SVs}\alpha_i y_i K(\mathbf{x}_i,\mathbf{x})+b\right) \tag{4}$$

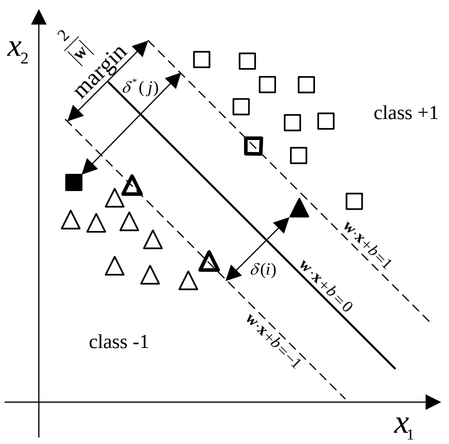

Fig. 2. Graphical description of the binary data classification using SVM model with slack variables.

2.2 SVMs for Regression Analysis

As stated above, after the introduction of Vapnik's ε-insensitive loss function [2,3], SVMs can be applied to a regression problem. In this case, SVMs can be termed support vector regression (SVR). The SVR model is used to resolve various problems such as a time series forecasting and a nonlinear regression.

The fundamental concept of SVR is to map the input data from the initial space into a kernel-induced higher dimensional space, and then to perform the linear regression analysis. In other words, nonlinear regression problems can be converted into linear regression problems in a characteristic space. The SVR function is generally expressed as follows:

$$\hat{y}=f(\boldsymbol{x})=\sum_{t=1}^{N}w_t\phi_t(\boldsymbol{x})+b=\boldsymbol{W}^T\boldsymbol{\Phi}(\boldsymbol{x})+b \tag{5}$$

where $\phi_t(\boldsymbol{x})$ means a characteristic that is nonlinearly changed from the input space $\boldsymbol{x}(t)$, $\boldsymbol{W}=\left[w_1 \quad w_2 \quad \cdots \quad w_N\right]^T$, and $\boldsymbol{\Phi}=\left[\phi_1 \quad \phi_2 \quad \cdots \quad \phi_N\right]^T$.

The following regularized risk function using ε-insensitive loss function has to be minimized to compute vector $\boldsymbol{W}$ and bias and acquire the optimal regression function [12].

$$R(\boldsymbol{W})=\frac{1}{2}\boldsymbol{W}^T\boldsymbol{W}+\mu\sum_{t=1}^{N}\left|f(\boldsymbol{x}(t))-y(t)\right|_\varepsilon \tag{6}$$

where

$$\left|f(\boldsymbol{x}(t))-y(t)\right|_\varepsilon=\begin{cases}0 & \text{if } \left|f(\boldsymbol{x}(t))-y(t)\right|<\varepsilon \\ \left|f(\boldsymbol{x}(t))-y(t)\right|-\varepsilon & \text{otherwise}\end{cases} \tag{7}$$

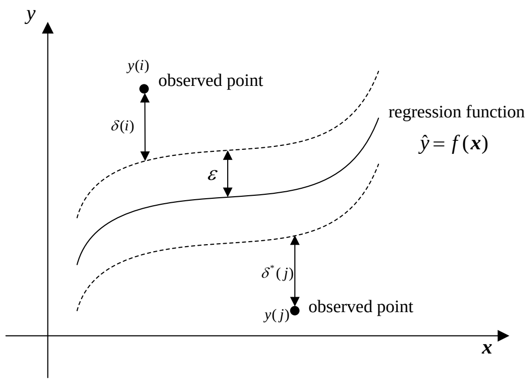

Fig. 3. Graphical description of the insensitive ε-tube with slack variables.

The ε-insensitive loss function in Eq. (6) is used for estimation stabilization (refer to Fig. 3) and this term is determined according to Eq. (7). Additionally, the parameter μ is a generalization parameter, which has a role to determine the trade-off between the number of support vectors and noisy data. ε and μ as user-specified parameters are related to generalization performance and overfitting.

Lastly, the optimal SVR function using the slack variables and kernel function becomes:

$$\hat{y} = f(\mathbf{x}) = \sum_{t=1}^{N}\left(\alpha_t - \alpha_t^*\right)K\left(\mathbf{x}, \mathbf{x}(t)\right) + b \qquad (8)$$

2.3 SVMs with Multiple Modules

In an effort to acquire the optimal performance of SVMs, the study for multiple connection of modules that comprised of the entire calculation process of aforementioned SVC or SVR models was carried out. In this paper, SVMs with multiple modules can be checked in Fig. 4. It consists of more than two modules and has a structure connected in series. The authors called this as cascaded structure and developed the cascaded support vector regression (CSVR) model to apply to regression analysis [8].

The CSVR method makes it possible to show the good performance by transferring the optimal output from a preceding module to the next module. Unfortunately, excessive increase in SVR modules may induce the overfitting problem in CSVR model. Therefore, it is necessary to prevent the overfitting problem and guarantee the best performance through proper generalization.

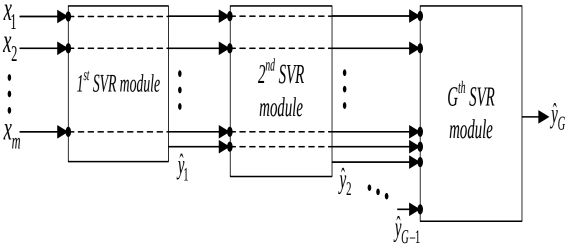

Fig. 4. SVR model in the cascaded structure [8].

3. Optimization of SVMs

It is obvious that the optimization of the developed SVC or SVR models is essential to show the best performance. The optimization of the SVM models can be achieved by selecting and learning from the informative data, and solving the overfitting problem.

3.1 Solving the Overfitting Problem

For the optimization of the SVM models, the data used for the machine learning methods are usually divided into learning data, validation data, and test data. Among them, the validation data are used to measure the generalization performance of estimation models and are associated with solving the overfitting problem.

The two methods for SVM optimization was proposed and used in previous studies. First, it was cross-validation using the validation data set as a scale of the generalization, which has a role to halt training when the generalization is getting worse. Secondly, a genetic algorithm such as selection, crossover, and mutation, and the fitness function were used to minimize the errors for a data set

3.2 Data Selection

It is important that the informative data have to be selected to efficiently train the SVM models as well. Several studies used a subtractive clustering (SC) scheme [13] to collect the data with the highest potential. Fig. 5 indicates an example of data selection using a SC scheme.

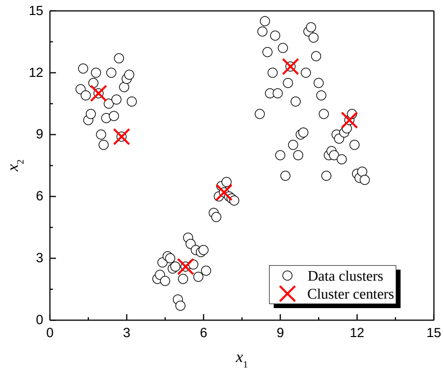

Fig. 5. An illustration of the informative data selection using SC scheme.

4. Application of SVMs

Several representative studies using these SVM models applied for identification, classification, and regression are described in this section.

4.1 Support Vector Classification

Major transients such as LOCAs where the break positions are hot-leg, cold-leg, and steam generator tube (SGT), total loss of feedwater (TLOFW), main steam line break (MSLB), feedwater line break (FWLB), and station blackout (SBO) of NPPs were identified using the integrated values of sensor signals [4]. Fig. 6 shows the identification of transients of NPPs using the SVC model

4.2 Support Vector Regression

The representative studies using the proposed SVR model are to estimate golden time for accident recovery [5], PPF [6], DNBR [7], LOCA break size [8], and cutter wear of a milling machine. Fig. 7 is a graph showing the comparison of estimated cutter wear versus actual cutter wear using the sensor signals such as force, acoustic emission, and vibration using the SVR model. The high accuracy of estimation can be checked.

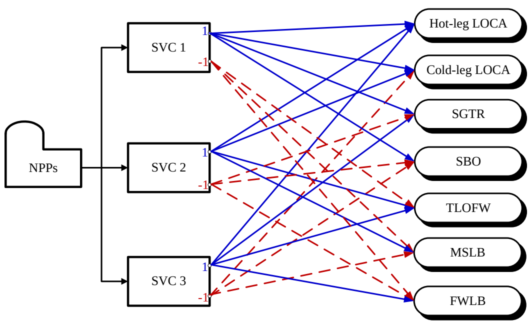

Fig. 6. Identification of transients of NPPs using the SVC model.

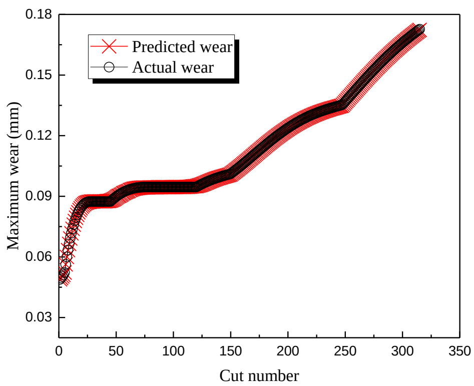

Fig. 7. Comparison of estimated cutter wear versus actual cutter wear of a milling machine.

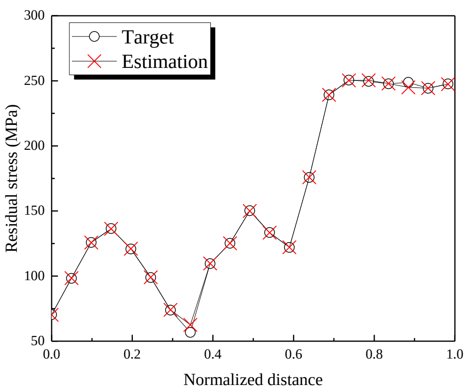

Fig. 8. Comparison of residual stress in welding of dissimilar metals.

4.3 Cascaded Support Vector Regression

The residual stress in welding of dissimilar metals were estimated using the CSVR model [8]. Likewise, the CSVR model shows outstanding estimation accuracy (refer to Fig. 8).

5. Conclusions

SVM is one of the representative machine learning method, which has been used for the studies to improve the safety of NPPs. To be specific, the SVM method was used for a classification problem such as identification of the transients of NPPs and regression problems such as estimation of golden time for accident recovery, PPF, DNBR, and cutter wear, which shows good performance and the applicability. Through application of these verified SVM models, it will be possible to enable to improve NPP safety, minimize human error, keep the integrity of internal equipment, and economically maintain the plants in the future.

REFERENCES

[1] C. Cortes and V. Vapnik, Suppot Vector Networks, Machine Learning, Vol.20, pp. 273-297, 1995.
[2] V. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, 1995.
[3] V. Vapnik, S. Golowich, and A. Smola, Support Vector Method for Function Approximation, Regression Estimation, MIT Press, Cambridge, Massachusetts, 1997.
[4] Y. D. Koo, J. H. Back, and M. G. Na, Identification of NPP Transients Using Artificial Intelligence, Proceedings of Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies(NPIC&HMIT 2017), Jun.11-15, 2017, San Francisco, CA.
[5] K. H. Yoo, J. H. Back, M. G. Na, J. H. Kim, S. Hur, and C. H. Kim, Prediction of Golden Time Using SVR for Recovering SIS under Severe Accidents, Annals of Nuclear Energy, Vol.94, pp. 102-108, 2016.
[6] I. H. Bae, M. G Na, Y. J. Lee, and G. C. Park, Calculation of the Power Peaking Factor in a Nuclear Reactor Using Support Vector Regression Models, Annals of Nuclear Energy, Vol.35, No.12, pp. 2200-2205, 2008.
[7] D. S. Kim, S. W. Lee, and M. G. Na, Prediction of Axial DNBR Distribution in a Hot Fuel Rod Using Support Vector Regression Models, IEEE Trans. Nucl. Sci., Vol.58, No.4, pp. 2084-2090, 2011.
[8] Y. D. Koo, K. H. Yoo, and M. G. Na, Estimation of Residual Stress in Welding of Dissimilar Metals at Nuclear Power Plants Using Cascaded Support Vector Regression, Nucl. Eng. Technol., Vol.49, No.4, pp. 817-824, 2017.
[9] M. G. Na, W. S. Park, and D. H. Lim, Detection and Diagnosis of Loss of Coolant Accidents Using Support Vector Machines, IEEE Trans. Nucl. Sci., Vol.55, No.1, 2008.
[10] S. R. Gunn, Support Vector Machines for Classification and Regression, Technical Report, University of Southampton, 1998.
[11] V. Vapnik, Statistical Learning Theory, John Wiley & Sons, New York, 1998.
[12] V. Kecman, Learning and soft computing, MIT Press, Cambridge, Massachusetts, 2001.
[13] S. L. Chiu, Fuzzy Model Identification Based on Cluster Estimation, J. Intell. Fuzzy Syst., Vol.2, pp. 267-278, 1994.