Journal of Robotics, Networking and Artificial Life
Vol. 8(1); June (2021), pp. 41–46
DOI: https://doi.org/10.2991/jrnal.k.210521.010; ISSN 2405-9021; eISSN 2352-6386
https://www.atlantis-press.com/journals/jrnal

Research Article

Anomaly Detection Using Support Vector Machines for Time Series Data

Umaporn Yokkampon[1,*], Sakmongkon Chumkamon[1], Abbe Mowshowitz[2], Ryusuke Fujisawa[1], Eiji Hayashi[1]

[1]*Graduate School of Computer Science and Systems Engineering, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan*
[2]*Department of Computer Science, The City College of New York, 160 Convent Avenue, New York, NY 10031, USA*

ARTICLE INFO

Article History

Received 26 November 2020
Accepted 12 April 2021

Keywords

Anomaly detection
support vector machine
data mining
factory automation

ABSTRACT

Analysis of large data sets is increasingly important in business and scientific research. One of the challenges in such analysis stems from uncertainty in data, which can produce anomalous results. This paper proposes a method for detecting an anomaly in time series data using a Support Vector Machine (SVM). Three different kernels of the SVM are analyzed to predict anomalies in the UCR time series benchmark data sets. Comparison of the three kernels shows that the defined parameter values of the Radial Basis Function (RBF) kernel are critical for improving the validity and accuracy in anomaly detection. Our results show that the RBF kernel of the SVM can be used to advantage in detecting anomalies.

1. INTRODUCTION

Research on anomaly detection is of great interest in machine learning and data mining. Detecting anomalies or finding outliers involves identifying abnormal or inconsistent patterns in a dataset. Abnormal data often results from unauthorized activity. Credit card fraud offers a well-known example. Transactions with a stolen or fake credit card can produce suspicious data. A fake card can be made by copying information from an authorized card and using it to create a new unauthorized one. Data such as personal identifying information may be obtained through phishing or from employees who work in credit card companies [1]. Another source of abnormal data may derive from unauthorized intrusions in networks. Abnormal traffic or user actions are common signs of intrusions, which may occasion breaches of sensitive or confidential data. Intrusions may also cause sensor networks to generate erroneous data. When a sensor malfunctions, it is unable to capture data correctly and thus may produce anomalies. Abnormal changes in data sources may also result in anomalies [2].

Anomaly detection typically uses data mining and machine learning methods for detecting abnormal activities in systems. Many anomaly detection techniques have been developed, including Support Vector Machines (SVM), which can solve classification and regression problems. The performance of SVM depends on the selection of kernel function and kernel parameters. The selection quality of SVM parameters and kernel functions has an effect on learning and generation performance. Appropriate kernel function and associated parameters should be selected to obtain optimal classification performance. When an appropriate kernel function and parameters are selected, the prediction error of SVM can be minimized.

This paper reports on application of the support vector machine method to eight real world time series data sets to detect anomalies using three different kernels for analysis and prediction. In addition, SVM kernels are compared for effectiveness based on Area under the Curve (AUC), Precision, Recall, *F*1-Score, Specificity, and Jaccard index criteria.

2. SUPPORT VECTOR MACHINE ALGORITHM

The SVM algorithm's goal is to create the best line or decision boundary that can decompose an n-dimensional space into sets supporting categorization of new data points. Hyperplanes define the boundaries in this space.

In our proposed method, we used SVM provided by Scikit-learn to detect the anomaly in time series data.

For a given dataset x with a number i of training data, SVM finds the maximum margin hyperplane separating different classes of data [3]:

$$x = (x_i, y_i),\ x_i \in \mathbb{R}^p,\ y_i \in \{-1,1\},\ \forall\ i = 1, 2, ..., N \tag{1}$$

where x_i is the p-dimensional input vector, y_i is the output value (1 or −1) and "·" is the dot product which has a formula form by $x \cdot y = \sum x_i y_i$. A decision vector separating two classes is given by:

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

*Corresponding author. Email: may@mmcs.mse.kyutech.ac.jp

where w^T is the optimal weighting vector and b is the bias. The margins of linearly separable training data can be defined as:

$$w^T \cdot x + b = 1 \text{ and } w^T \cdot x + b = -1 \tag{3}$$

The distance between the margins can be defined by $2/\|w^T\|$. Therefore, minimizing $\|w^T\|$ is an objective function. In practice, it is not easy to linearly decompose the training dataset. Let C be the regularization parameter that defines the separation of two classes and the error when using a training dataset. The hyperplane is determined by minimizing:

$$C\sum_{i=1}^{N}\varepsilon_i + \frac{1}{2}\|w\|^2 \tag{4}$$

with constraints $t_i y(x_i) \geq 1-\varepsilon_i$, $i = 1, \ldots, N$ where t_i is the target value and ε_i is the set of slack variables.

Instead of using a minimization model (4), the problem may be formulated using Lagrangian dual multipliers α as:

$$\max \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i, x_j) \tag{5}$$

subject to:

$$0 \leq \alpha_i \leq C \;\; \forall \;\; i = 1, 2, \ldots, n \text{ and } \sum_{i=1}^{n}\alpha_i y_i = 0 \tag{6}$$

Kernel trick can be applied to reduce the complexity of the optimization problem.

Support vector machine with nonlinear kernel has an objective function form as follows:

$$\max \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j k(x_i, x_j) \tag{7}$$

3. KERNELS

In machine learning, kernel methods are a popular class for a variety of tasks. Kernel methods can generate the model complex data through the kernel trick, which is an important feature [4].

In this paper, two types of kernel functions are chosen and evaluated, namely the linear and Radial Basis Function (RBF). The mathematical formula for the said functions are as follows:

3.1. Linear Kernel

$$k(x_i, x_j) = (x_i, x_j) \tag{8}$$

3.2. Radial Basis Function Kernel

$$k(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{(2\sigma^2)}\right) \tag{9}$$

$K(x_i, x_j) = x_i \cdot x_j$ is the kernel function, which is generalized dot products. "·" denotes the dot product. The two kernels have their own advantages and limitations. The linear kernel offers ease of performance as well as an ability to deal with small and linearly separable samples. The RBF kernel, on the other hand, is known as a good mapping function since it can be used for all kinds of samples, small or large with both high and low dimensions [5]. The SVM performance for each kernel will be evaluated in this study to determine the optimal kernel.

4. EXPERIMENTS

This section introduces the data sets and the evaluation metric used. We have compared the three kernels and evaluated their effectiveness for anomaly detection in SVMs.

4.1. Data Sets

Time series data obtained from UCR public data set [6] were used to evaluate effectiveness. Table 1 shows the details of the datasets. All datasets are presented in time series form, and every data point is manually labeled. For all datasets, we designated the minority class as an anomaly class. Twenty percent of the data was used for testing.

4.2. Evaluation Metrics

The accuracy of an anomaly detection method is evaluated using the AUC of the Receiver Operating Characteristic (ROC), Precision (Pre), Recall (Rec), $F1$-Score, Specificity, and Jaccard index, defined as follows (Figure 1):

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{10}$$

Table 1 | Summary of the datasets

Datasets	Length	Number of instances	Anomaly ratio
ItalyPowerDemand	24	1096	0.49
Wafer	152	7164	0.11
SonyAIBORobotSurface2	65	980	0.38
ECGFiveDays	136	884	0.50
TwoLeadECG	82	1162	0.50
MoteStrain	84	1272	0.46
Herring	512	128	0.40
Strawberry	235	983	0.36

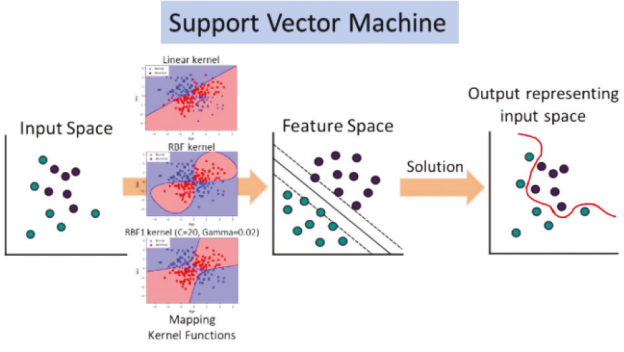


Figure 1 | Structure of anomaly detection in time series data used SVM. We used eight time series data sets processed by SVM, and three different kernels based on AUC, Precision, Recall, $F1$-Score, Specificity, and Jaccard index criteria.

$$\text{Recall} = \text{Sensitivity} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{11}$$

$$F1\text{-Score} = 2 \times \frac{\text{Pre} \times \text{Rec}}{\text{Pre} + \text{Rec}} \tag{12}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN}+\text{FP}} \tag{13}$$

$$\text{Jaccard} = \frac{\text{TP}}{\text{TP}+\text{FP}+\text{FN}} \tag{14}$$

where TP is the accurately detected abnormal, FP is the false detected abnormal, TN is the accurately assigned normal, and FN is the false assigned normal.

5. RESULTS AND DISCUSSION

The efficiency of the following three SVM kernels are compared:

1. Linear Kernel

2. RBF Kernel (Default parameters value)

3. RBF1 Kernel (We define the parameters $C = 20$, $\gamma = 0.02$)

We performed experiments on accuracy of analysis and prediction of anomalies for eight time series data sets using the three different SVM kernels. Accuracy of analysis and prediction can be measured by the AUC as shown in the ROC in Figures 2–9. The blue line is the Linear Kernel of SVM, the orange line is the RBF kernel, and the green line is the RBF1 kernel.

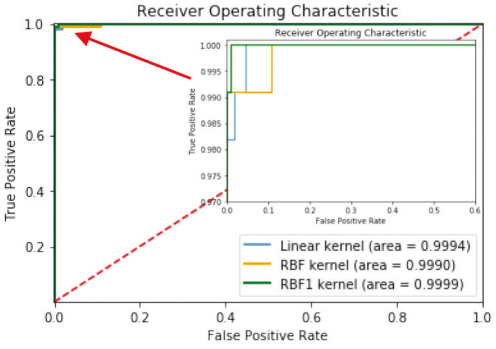


Figure 2 | The kernel performance comparison of Linear, RBF and RBF1 for testing ItalyPowerDemand dataset using ROC.

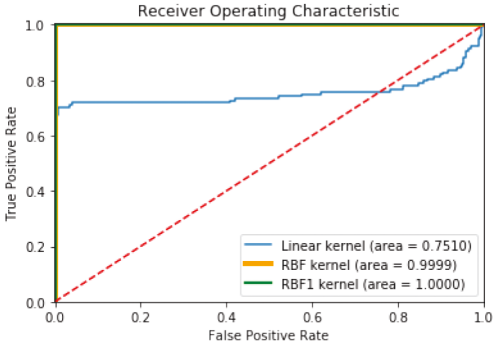


Figure 3 | The kernel performance comparison of Linear, RBF and RBF1 for testing Wafer dataset using ROC.

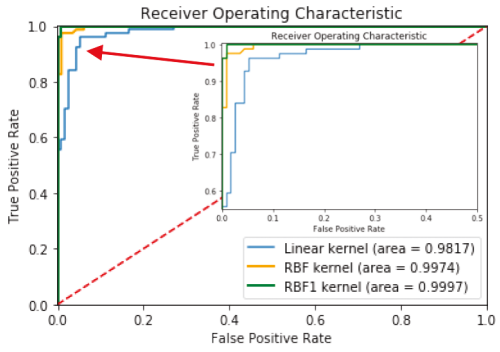


Figure 4 | The kernel performance comparison of Linear, RBF and RBF1 for testing SonyAIBORobotSurface2 dataset using ROC.

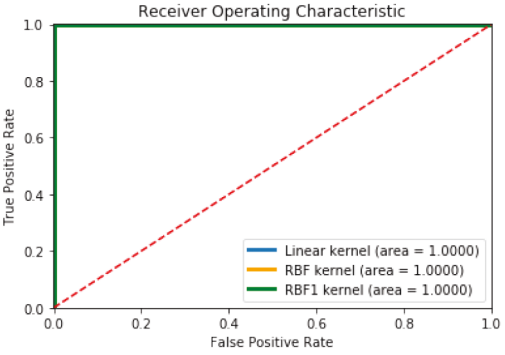


Figure 5 | The kernel performance comparison of Linear, RBF and RBF1 for testing ECGFiveDays dataset using ROC.

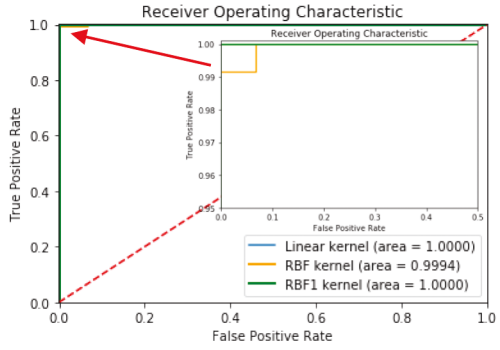


Figure 6 | The kernel performance comparison of Linear, RBF and RBF1 for testing TwoLeadECG dataset using ROC.

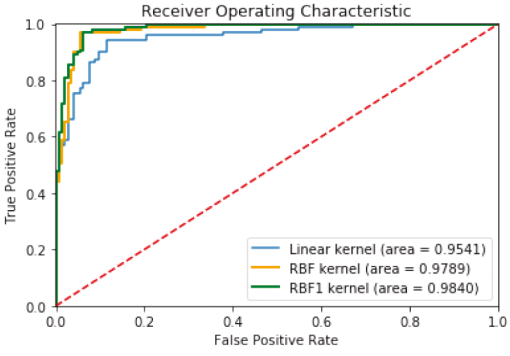


Figure 7 | The kernel performance comparison of Linear, RBF and RBF1 for testing MoteStrain dataset using ROC.

Figure 2 shows that the RBF1 kernel is slightly more efficient than the Linear and RBF kernels for the ItalyPowerDemand data set. However, all three kernels yield almost 100% accuracy.

Figure 3 shows that the RBF and RBF1 kernels are more efficient than the Linear kernel, and that the RBF kernel is almost 100% accurate. In particular, the RBF1 kernel gives a ROC value perfectly for the Wafer data set.

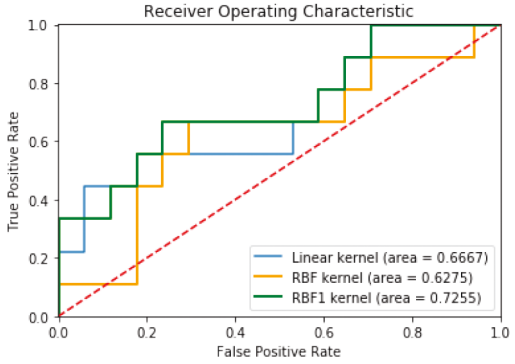


Figure 8 | The kernel performance comparison of Linear, RBF and RBF1 for testing Herring dataset using ROC.

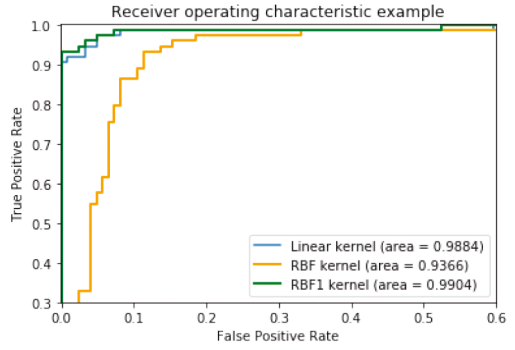


Figure 9 | The kernel performance comparison of Linear, RBF and RBF1 for testing Strawberry dataset using ROC.

Figure 4 reveals that RBF1 is slightly more efficient than the linear and RBF kernels for the SonyAIBORobotSurface2 data set. In particular, the RBF and RBF1 kernels provide almost 100% accuracy.

Figure 5 shows that all three kernels give perfect ROC values for the ECGFiveDays data set.

Figure 6 shows that the RBF kernel is almost 100%. Linear and RBF1 kernels give nearly perfect ROC values for the TwoLeadECG data set.

Figure 7 shows that the RBF1 kernel to be slightly more efficient than the linear and RBF kernels for the MoteStrain data set.

Figure 8 reveals that the Linear kernel gives slightly more accurate ROC values than does the RBF kernel, but the RBF1 kernel is the most accurate for the Herring data set.

Finally, Figure 9 shows that the Linear kernel is more accurate than the RBF kernel. In particular, the RBF1 kernel is almost 100% for the Strawberry data set.

Table 2 shows the summary of anomaly detection results and comparisons. The results show that SVM with RBF1 kernel gives the highest accuracy, specificity, Jaccard index, and *F*1-Score on all aspects and data sets, except for the Herring data set, for which the RBF kernel gives the highest of *F*1-Score and Jaccard. All three kernels gave perfect results for AUC, Precision, Recall, *F*1-Score, Specification, and Jaccard index on the ECGFiveDays data set.

We compared our results to the latest research in "Time Series Anomaly Detection with Variational Autoencoders [7]", which used a different method. There are six data sets utilized in our method, namely, ItalyPowerDemand, Wafer, ECGFiveDays, TwoLeadECG, MoteStrain, and Herring. Table 3 shows the AUC results of the latest research. These results demonstrate that our proposed

Table 2 | Summary of the kernel performance comparison of Linear, RBF, and RBF1

Datasets	Kernel: Linear				Kernel: RBF (Default)				Kernel: RBF ($C = 20$, $\gamma = 0.02$)			
	AUC	Precision	Recall	*F*1-Score	AUC	Precision	Recall	*F*1-Score	AUC	Precision	Recall	*F*1-Score
ItalyPowerDemand	0.9994	0.9818	0.9818	0.9818	0.9990	0.9910	1.0000	**0.9955**	**0.9999**	0.9910	1.0000	**0.9955**
Wafer	0.7510	0.9687	0.9946	0.9815	0.9999	1.0000	0.9969	0.9985	**1.0000**	1.0000	0.9985	**0.9992**
SonyAIBORobotSurface2	0.9819	0.9646	0.9478	0.9561	0.9974	0.9910	0.9565	0.9735	**0.9997**	1.0000	0.9826	**0.9912**
ECGFiveDays	1.0000	1.0000	1.0000	1.0000	**1.0000**	1.0000	1.0000	1.0000	**1.0000**	1.0000	1.0000	**1.0000**
TwoLeadECG	1.0000	0.9915	1.0000	0.9957	0.9994	0.9915	1.0000	0.9957	**1.0000**	1.0000	1.0000	**1.0000**
MoteStrain	0.9540	0.9552	0.8767	0.9143	0.9789	0.9783	0.9247	0.9507	**0.9840**	0.9716	0.9384	**0.9547**
Herring	0.6667	0.7647	0.7647	0.7647	0.6275	0.6538	1.0000	**0.7907**	**0.7255**	0.7692	0.5882	0.6667
Strawberry	0.9884	0.9908	0.8710	0.9270	0.9366	0.6327	1.0000	0.7750	**0.9904**	0.9911	0.8952	**0.9407**

Datasets	Kernel: Linear		Kernel: RBF (Default)		Kernel: RBF ($C = 20$, $\gamma = 0.02$)	
	Specificity	Jaccard	Specificity	Jaccard	Specificity	Jaccard
ItalyPowerDemand	0.9818	0.9643	**0.9909**	**0.9910**	**0.9909**	**0.9910**
Wafer	0.6719	0.9636	1.0000	0.9969	**1.0000**	**0.9985**
SonyAIBORobotSurface2	0.9506	0.9160	0.9877	0.9483	**1.0000**	**0.9826**
ECGFiveDays	1.0000	1.0000	1.0000	1.0000	**1.0000**	**1.0000**
TwoLeadECG	0.9914	0.9915	0.9914	0.9915	**1.0000**	**1.0000**
MoteStrain	0.9450	0.8421	**0.9725**	0.9060	0.9633	**0.9133**
Herring	0.5556	0.6190	0.0000	**0.6538**	**0.6667**	0.5000
Strawberry	0.9863	0.8640	0.0137	0.6326	**0.9863**	**0.8880**

Table 3 | Comparing AUC of VAE results from latest research

Datasets	OUR	ANOGAN	ALAD	MLP-VAE	IF
ItalyPowerDemand	0.761	0.516	0.538	0.768	0.763
Wafer	**0.965**	0.558	0.587	0.790	0.847
ECGFiveDays	**0.970**	0.970	0.694	0.910	0.678
TwoLeadECG	**0.891**	0.554	0.515	0.731	0.760
MoteStrain	**0.840**	0.746	0.504	0.750	0.762
Arrhythmia	**0.758**	0.576	0.515	0.747	0.530
KDD99	**0.958**	0.887	0.950	0.622	0.929
GunPointAgeSpan	**0.881**	0.515	0.547	0.821	0.612
ToeSegmentation2	**0.846**	0.547	0.544	0.816	0.787
Herring	0.659	0.488	0.569	0.627	0.698

method achieves superior results compared to the approach taken in Zhang and Chen [7] for all six datasets.

This shows that the RBF kernel with parameter values $C = 20$, $\gamma = 0.02$ exhibits good performance in anomaly detection for time series data.

6. CONCLUSION

In this paper, we presented an analysis of anomaly detection in time series data using SVM with three different kernels, namely, Linear, RBF and RBF1. We evaluated the accuracy of anomaly detection methods based on AUC, Precision, Recall, *F*1-Score, Specificity, and Jaccard index criteria. The evaluation results show that the kernel with defined parameters can improve accuracy on all aspects and data sets. This application of the SVM method, with the RBF kernel, can be efficient for detecting anomalies in time series data. The results for data set ECGFiveDays show 100% accuracy with all three kernels, and the results for the TwoLeadECG show almost 100% with all three kernels. Moreover, the results indicate a high degree of accuracy for the three kernels on all the data sets, perhaps because our data was trained in supervised conditions. Since we train the machine using data that is well labeled, and the algorithms learn to predict output from the input data. It means some data is already tagged with the correct answer.

In the future, we intend to implement the variational autoencoder method for detecting and predicting anomalies in time series and spectrum data to compare it with the autoencoder method.

CONFLICTS OF INTEREST

The authors declare they have no conflicts of interest.

ACKNOWLEDGMENT

This work was supported by the Religion revitalization project by the Japanese Government.

REFERENCES

[1] E. Hormozi, M. Kazem Akbari, H. Hormozi, M.S. Javan, Accuracy evaluation of a credit card fraud detection system on Hadoop MapReduce, The 5th Conference on Information and Knowledge Technology, IEEE, Shiraz, Iran, 2013, pp. 35–39.

[2] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: a survey, ACM Comput. Surv. 41 (2009), 15.

[3] P. Batta, M. Singh, Z. Li, Q. Ding, L. Trajković, Evaluation of support vector machine kernels for detecting network anomalies, 2018 IEEE International Symposium on Circuits and Systems (ISCAS), IEEE, Florence, Italy, 2018, pp. 1–4.

[4] B. Schölkopf, A.J. Smola, Support vector machines and kernel algorithms, in: P. Armitage, T. Colton (Eds.), Encyclopedia of Biostatistics, John Wiley & Sons, New York, USA, 2005, pp. 5328–5335.

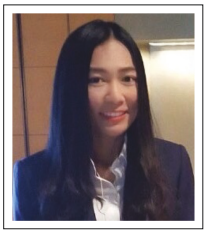
[5] W.F. Cao, L. Li, X.L. Lv, Kernel function characteristic analysis based on support vector machine in face recognition, 2007 International Conference on Machine Learning and Cybernetics, IEEE, Hong Kong, China, 2007, pp. 2869–2873.

[6] H.A. Dau, E. Keogh, K. Kamgar, C.C.M. Yeh, Y. Zhu, S. Gharghabi, et al., The UCR time series classification archive, 2019. Available from: https://www.cs.ucr.edu/~eamonn/time_series_data_2018/.

[7] C. Zhang, Y. Chen, Time series anomaly detection with variational autoencoders, CoRR, abs/1907.01702, 2019.

AUTHORS INTRODUCTION

Ms. Umaporn Yokkampon



She received Master of Science degree in Applied Statistics from King Mongkut's University of Technology North Bangkok in 2019. Presently, she is a doctoral student in the department of Computer Science and Systems Engineering at Kyushu Institute of Technology. Her research interests include data mining, anomaly detection, and time series analysis.

Dr. Sakmongkon Chumkamon

He received Doctor of Engineering degree from Kyushu Institute of Technology in 2017. He was a postdoctoral researcher at Guangdong University of Technology in 2017–2019. Presently, he is a postdoctoral researcher in Kyushu Institute of Technology since 2019. His research interests include factory automation robots and social robots.

Prof. Abbe Mowshowitz

He received the PhD degree from University of Michigan in 1967. He has been a professor of computer science at the City College of New York and member of the doctoral faculty at the Graduate Center of the City University of New York since 1984. His current research interests lie in two areas are organizational and managerial issues in computing, and network science. In addition to teaching and research, he has acted as consultant on the uses and impacts of information technology (especially computer networks) to a wide range of public and private organizations in North America and Europe.

Assoc. Prof. Ryusuke Fujisawa

He is an associate professor in the department of Intelligent and Control Systems at Kyushu Institute of Technology. He received the PhD (Dr. Eng.) degree from The University of Electro-Communications in 2009. His research interests include Swarm intelligence, Swarm robotics and Distributed autonomous system. He is a member of Japan Society of Civil Engineers (JSCE), Information Processing Society of Japan (IPSJ), The Society of Instrument and Control Engineers (SICE) and The Robotics Society of Japan (RSJ).

Prof. Eiji Hayashi

He is a professor in the department of Intelligent and Control Systems at Kyushu Institute of Technology. He received the PhD (Dr. Eng.) degree from Waseda University in 1996. His research interests include intelligent mechanics, mechanical systems and perceptual information processing. He is a member of The Institute of Electrical and Electronics Engineers (IEEE), and The Japan Society of Mechanical Engineers (JSME).