



Published in final edited form as:

*Int Symp Process Syst Eng.* 2018 ; 44: 2077–2082. doi:10.1016/B978-0-444-64241-7.50341-4.

## Simultaneous Fault Detection and Identification in Continuous Processes via nonlinear Support Vector Machine based Feature Selection

Melis One<sup>a,b</sup>, Chris A. Kieslich<sup>c</sup>, Yannis A. Guzman<sup>a,b,d</sup>, and Efstratios N. Pistikopoulos<sup>a,b</sup>

<sup>a</sup>Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX, 77843, USA

<sup>b</sup>Texas A&M Energy Institute, Texas A&M University, College Station, TX, 77843, USA

<sup>c</sup>Coulter Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, 30332, USA

<sup>d</sup>Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ, 08544, USA

### Abstract

Rapid detection and identification of process faults in industrial applications is crucial to sustain a safe and profitable operation. Today, the advances in sensor technologies have facilitated large amounts of chemical process data collection in real time which subsequently broadened the use of data-driven process monitoring techniques via machine learning and multivariate statistical analysis. One of the well-known machine learning techniques is Support Vector Machines (SVM) which allows the use of high dimensional feature sets for learning problems such as classification and regression. In this paper, we present the application of a novel nonlinear (kernel-dependent) SVM-based feature selection algorithm to process monitoring and fault detection of continuous processes. The developed methodology is derived from sensitivity analysis of the dual SVM objective and utilizes existing and novel greedy algorithms to rank features that also guides fault diagnosis. Specifically, we train fault-specific two-class SVM models to detect faulty operations, while using the feature selection algorithm to improve the accuracy of the fault detection models and perform fault diagnosis. We present results for the Tennessee Eastman process as a case study and compare our approach to existing approaches for fault detection, diagnosis and identification.

### Keywords

Fault Detection and Identification; Process Monitoring; Data-driven Modeling; Feature Selection; Support Vector Machines

## 1. Introduction

Maintaining a safe operation while minimizing the losses in productivity is one of the major goals in chemical processing. Timely detection, diagnosis and identification of faults occurring during the operation is critical and necessary, and this is continuously encouraging researchers to develop novel process monitoring techniques. Today, with the advances in automation and sensor technologies as well as novel smart manufacturing frameworks, real time process data acquisition has become effortless and created an immense opportunity for manufacturing and process industries by enabling data-driven real time decision making through it.

A fault is defined as an abnormal process behaviour where controllers lack the capability of reversing it. Faults can occur due to mechanical reasons such as equipment failure and/or wear as a result of equipment aging or process-based reasons stemming from severe process disturbances (Chiang et al., 2001). Once a fault is detected during an operation, revealing the root-cause by identifying the key process variables is of utmost importance for rapid counteraction to avoid abnormal event progression (Venkatasubramanian et al., 2003). The most prevalent data-driven process monitoring techniques include Principal Component Analysis (PCA), Dynamic PCA (DPCA), Partial Least Squares (PLS), Independent Component Analysis (ICA), and Fischer Discriminant Analysis (FDA) as dimensionality reduction step followed by  $Q$  and  $T^2$  statistics, contribution plots, and discriminant analysis (Qin, 2012). The aim of dimensionality reduction is to ensure robustness of the analysis. Dimensionality reduction methods are classified into (a) feature extraction, and (b) feature selection. The dimensionality reduction methods adopted within the listed techniques are based on feature extraction, which project input process data onto another space and alter the original representation of input data, thus may cause loss of information. On the other hand, feature selection is the process of selecting the most informative and relevant original features (e.g. process variables) characterizing the system. Therefore, there is a prominent need for novel data-driven process monitoring methods incorporating feature selection techniques for dimensionality reduction.

In this work, we modify a well-known, powerful machine learning algorithm formulation, Support Vector Machines (SVM) (Cortes and Vapnik, 1995), for simultaneous modelling and feature selection. Then, we present the application of the nonlinear (Kernel-dependent) SVM-based feature selection algorithm to process monitoring of continuous processes. Previous studies have used SVM for fault detection in chemical processes (Mahadevan and Shah, 2009; Yin et al., 2014; Xiao et al., 2016). Here, fault detection is achieved through two-class SVM models, where the feature selection algorithm further improves the model accuracy and also reveals diagnosis of the detected fault. We test the performance of the proposed data-driven framework through the 21 faults introduced in the Tennessee Eastman process data and provide comparisons to existing approaches. The presented methodology can be implemented as an online decision support tool for continuous process monitoring purposes.

## 2. Nonlinear Support Vector Machine-based Feature Selection Algorithm

Here, we define a supervised learning problem with  $l$  training instances, where  $x_j \in R^n$ . Indices  $i, j = 1, 2, \dots, l$  correspond to instances, whereas indices  $k, k' = 1, 2, \dots, n$  correspond to input data features. Accordingly, the input data has become  $x_i = (x_{i1}, x_{i2}, \dots, x_{ik}, \dots, x_{in})^T$ . From process monitoring perspective, instances represent distinct continuous operations while features are the collected process measurements from distinct variables. We formulate the fault detection problem in a classification setting with  $C$ -parameterized SVM ( $C$ -SVM) with hinge loss,  $l_2$ -norm penalty, and linear Kernel as follows:

$$\begin{aligned} \min_{w, b, \xi} & \frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^l \xi_i \\ \text{s.t.} & \\ & y_i(w \cdot x_i + b) \geq 1 - \xi_i \quad i = 1, 2, \dots, l \\ & \xi_i \geq 0 \quad i = 1, 2, \dots, l \end{aligned} \quad (1)$$

where  $y_i \in \{-1, 1\}$  denotes the class label of instance  $i$ . Eq.(1) is a convex nonlinear problem satisfying a first-order qualification, hence strong duality holds. When Eq.(1) is solved to global optimality, resulted optimal solution  $(w^*, b^*, \xi^*)$  is then used to determine the linear decision function  $f(x) = w^* \cdot x + b^*$ , which yields the sign, thus class, of a new instance  $x$ .

One of the major advantages of exploiting Support Vector Machines in modelling is to be able to use Kernel functions,  $K(x_i, x_j)$ . These functions implicitly map the input feature space, where data is nonlinearly separable, onto a different, possibly higher feature space, where the data become linearly separable. Since most of the chemical processes are nonlinear in nature, we adopt nonlinear (Kernel-dependent)  $C$ -SVM formulation, where Kernel functions are introduced in the Lagrange dual formulation of Eq.(1). Next, we introduce binary variables  $z \in \{0, 1\}^n$  in Eq.(2), which control elimination or selection of feature  $k$ , and aim to minimize the number of selected features while maximizing model accuracy via traditional  $C$ -SVM formulation:

$$\begin{aligned} \min_{\alpha, z} & \max_{\alpha} \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(x_i \circ z, x_j \circ z) \\ \text{s.t.} & \\ & \sum_{i=1}^l \alpha_i y_i = 0 \\ & \alpha_i \in [0, C] \quad i = 1, 2, \dots, l \\ & \sum_k z_k = m \\ & z_k \in \{0, 1\} \quad k = 1, 2, \dots, n \end{aligned} \quad (2)$$

where  $\alpha_i$  are dual variables. In Eq.(2),  $\circ$  is the Hadamard product operator for componentwise multiplication, and  $m$  is the number of optimal subset of features, where the

aim is to attain highest  $C$ -SVM model accuracy with minimum number of input features. Since Eq.(2) is a challenging and impractical problem to solve to global optimality, we propose an algorithmic solution procedure, which utilizes the Lagrangian sensitivity of the objective function value of Eq.(2) with respect to  $z_k$  at  $(\alpha^*; z)$ , where  $z_k$  is treated as a fixed parameter. The procedure is iterative where in each step, the features are eliminated individually according to the following criterion:

$$crit_k = -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i^* \alpha_j^* y_i y_j \frac{\partial K(x_i \circ z, x_j \circ z)}{\partial z_k} \Bigg|_{z=1} \quad (3)$$

$$k_{worst} = arg_k \max crit_k$$

Eq.(3) delineates the nonlinear SVM-based feature selection algorithm adopted in this study. In particular, the algorithm is equivalent to well-known recursive feature elimination (RFE)-SVM classification algorithm when performing a linear classification that is using a linear Kernel. The presented algorithm (Kieslich et al., 2016) has been implemented in C++/Python environment using LibSVM library (Chang and Lin, 2011), and has been successfully utilized in our previous studies in bioinformatics setting.

### 3. Tennessee Eastman Process

The Tennessee Eastman process simulation, an extensively used benchmark study to evaluate and further develop process monitoring algorithms, was designed by the Eastman Chemical Company (Downs and Vogel, 1993). It is based on a real industrial process, in which the components, kinetics, and operating conditions have been modified for proprietary reasons (Chiang et al., 2001). The process includes five primary units: a reactor, condenser, compressor, separator, and a stripper. It describes the production of chemicals G and H from feedstocks A, C, D and E with byproduct F and inert compound B. The simulated process data used in this study is adopted from Chiang and Braatz, 2001, where the process contains 41 measured and 11 manipulated variables. The variables are sampled every 3 minutes and the dataset includes measurement samples from normal and 21 distinct faulty operations. For further information on the process and simulation, the interested reader can refer to Downs and Vogel, 1993 and Chiang et al., 2001.

### 4. Proposed Framework

In this study, we are building 21 fault-specific  $C$ -SVM binary classifiers for 21 different faults introduced in the process data. Thus, for each of the 21 model building, we combine data from normal and faulty operation from each of the 21 faults. The initial step in data-driven modelling is normalization of the input data. Normalization is performed on each of the 21 faulty-normal process data by subtracting the mean of measurements across the operations from each measurement and then dividing into the standard deviation of them.

#### 4.1. Offline Phase Model Building

The initial step in the offline phase is to create balanced train and test sets. Imbalanced data sets may significantly increase risk of overfitting problem. Therefore we create train and test sets from each 21 datasets with 100 runs of 5-fold cross-validation. Next, we build binary classifier models iteratively for each feature subset. In each iteration we (i) tune  $C$ -SVM parameters using train and test data sets with the active set of features (whole feature set in the first iteration), (ii) train  $C$ -SVM classifiers with Gaussian Radial Basis Function, where the class probabilities are smoothened via taking median of probabilities with a window size of 3, (iii) calculate Lagrangian sensitivity of dual  $C$ -SVM objective function with respect to the parameter  $z_k$  to obtain a rank list of features, (iv) eliminate the most redundant, "worst" feature satisfying Eq.(3) from data. This procedure is repeated until we are left with 1 feature in the train and test sets. The result of this framework yields 1092  $C$ -SVM classifiers, which includes 21 fault-specific  $C$ -SVM classifiers for each of 52 feature subsets.

Particularly, in the first step, we tune  $C$ -SVM hyperparameters  $C$  and  $\gamma$ . Tuning is done via grid search for all combinations between  $2^{-10}$  and  $2^{10}$ . Hyperparameter combination yielding the highest average testing accuracy and recall is chosen in each iteration of the framework described above. The final stage in the offline phase is the selection of the fault-specific end-models for fault detection. Among the all feature subsets, the  $C$ -SVM classifier yielding the highest fault detection rate and accuracy is chosen to be the fault-specific end-model.

#### 4.2. Online Phase Model Building

21 fault-specific end models that are built in offline phase are implemented in industrial setting to monitor the incoming online process data. When online process data is fed, models generate a binary answer for detection of each fault. If a fault is detected, alarm rises and the optimal set of features of the corresponding end-model yield the root-cause analysis of the detected fault instantaneously.

### 5. Results

We present the results of the proposed data-driven process monitoring framework on the Tennessee Eastman process below (Table 1), where we have adopted a five consecutive fault alarm policy to report a fault. The performance of the end-models is evaluated via (i) fault detection rate, (ii) accuracy, and (iii) latency.

Table 2 compares the performance of this study to the other existing data-driven process monitoring techniques (Mahadevan and Shah, 2009; Yin et al., 2014; Xiao et al., 2016) in terms of fault detection rate. This comparison demonstrates the power of the proposed data-driven process monitoring framework. Additionally, the average latency among the reported faults has been provided as 306.19, 145.58, 263.12, 151.00, and 98.50 min for PCA-T<sup>2</sup>, PCA-Q, DPCA-T<sup>2</sup>, DPCA-Q, and 1-class SVM respectively in Mahadevan and Shah, 2009; whereas it is significantly lower for the proposed framework (15.67 min for the Table 2 reported faults, and 13.86 min among all 21 faults).

## 6. Conclusions

In this study, we have presented a nonlinear SVM-based feature selection algorithm procedure and implemented it for data-driven process monitoring of continuous processes. The proposed framework establishes a promising decision support tool for online fault detection and identification. This research was funded by U.S. National Institute of Health (NIH) grant P42 ES027704.

## References

- Chiang LH, Russell EL, Braatz RD, 2001, Fault Detection and Diagnosis in Industrial Systems, Springer.
- Venkatasubramanian V, Rengaswamy R, Yin K, Kavuri SN, 2003, A Review of Process Fault Detection and Diagnosis Part I: Quantitative Model-based Methods, *Comp Chem Eng*, 27, 293–311.
- Qin SJ, 2012, Survey on Data-driven Industrial Process Monitoring and Diagnosis, *Annual Reviews in Control*, 36, 220–234.
- Cortes C and Vapnik V, 1995, Support Vector Networks, *Machine Learning*, 20, 273–297.
- Mahadevan S and Shah SL, 2009, Fault Detection and Diagnosis in Process Data Using One-class Support Vector Machines, *Journal of Process Control*, 19, 1627–1639.
- Yin S, Gao X, Karimi HR, Zhu X, 2014, Study on Support Vector Machine-Based Fault Detection in Tennessee Eastman Process, *Abstract and Applied Analysis*, Hindawi.
- Xiao Y, Wang H, Xu W, Zhou J, 2016, Robust one-class SVM for fault detection, *Chemometrics and Intelligent Laboratory Systems*, 151, 15–25.
- Kieslich CA, Tamamis P, Guzman YA, Onel M, Floudas CA, 2016, Highly Accurate Structure-based Prediction of HIV-1 Coreceptor Usage Suggests Intermolecular Interactions Driving Tropism, *PLoS One*, 11.
- Downs JJ and Vogel EF, 1993, A Plant-wide Industrial Process Control Problem, *Computers and Chemical Engineering*, 17, 245–255.

**Table 1.**

Performance of the Fault Specific End Models.

Fault	Optimal Feature Subset Size	Fault Detection Rate	Accuracy	Latency (min)
1	2	99.88	99.90	6
2	8	98.13	98.44	48
3	4	100.00	83.33	3
4	1	100.00	100.00	3
5	25	100.00	100.00	3
6	1	100.00	100.00	3
7	1	100.00	100.00	3
8	6	99.38	96.88	3
9	6	99.50	83.33	3
10	9	98.25	83.85	3
11	29	96.63	86.46	3
12	9	100.00	96.15	3
13	2	95.00	89.48	3
14	3	100.00	100.00	3
15	7	98.25	82.81	3
16	26	100.00	87.92	3
17	7	97.63	84.69	3
18	24	90.38	91.56	183
19	3	100.00	83.33	3
20	4	100.00	83.75	3
21	1	100.00	99.90	3

**Table 2.**

Comparison through Fault Detection Rate. Best results of Xiao et al., 2016 is adopted.

Ref.	Mahadevan and Shah, 2009					Xiao et al, 2016	Yin et al, 2014	This Study
Fault	PCA-T2	PCA-Q	DPCA-T2	DPCA-Q	1-class SVM	1-class SVM	2-class SVM	2-class SVM
1	99.20	99.80	99.40	99.50	99.80	99.50	99.50	<b>99.90</b>
2	98.00	<b>98.60</b>	98.10	98.50	<b>98.60</b>	98.30	98.12	98.10
4	4.40	96.20	6.10	<b>100.00</b>	99.60	47.40	99.88	<b>100.00</b>
5	22.50	25.40	24.20	25.20	<b>100.00</b>	45.20	90.75	<b>100.00</b>
6	98.90	<b>100.00</b>	98.70	<b>100.00</b>	<b>100.00</b>	99.20	60.13	<b>100.00</b>
7	91.50	<b>100.00</b>	84.10	<b>100.00</b>	<b>100.00</b>	70.10	98.91	<b>100.00</b>
8	96.60	97.60	97.20	97.50	97.90	97.40	96.00	<b>99.38</b>
10	33.40	34.10	42.00	33.50	87.60	68.00	81.00	<b>98.25</b>
11	20.60	64.40	19.90	80.70	69.80	65.80	80.25	<b>96.62</b>
12	97.10	97.50	99.00	97.60	99.90	98.80	97.75	<b>100.00</b>
13	94.00	<b>95.50</b>	95.10	95.10	<b>95.50</b>	95.00	92.50	95.00
14	84.20	<b>100.00</b>	93.90	<b>100.00</b>	<b>100.00</b>	93.90	91.00	<b>100.00</b>
16	16.60	24.50	21.70	29.20	89.80	73.10	89.38	<b>100.00</b>
17	74.10	89.20	76.00	94.70	95.30	75.20	81.63	<b>97.62</b>
18	88.70	89.90	88.90	90.00	90.00	89.30	89.50	<b>90.38</b>
19	0.40	12.70	0.70	24.70	83.90	43.60	85.88	<b>100.00</b>
20	29.90	45.00	35.60	51.00	90.00	69.00	80.50	<b>100.00</b>
21	26.40	43.00	35.60	44.20	52.80	59.40	-	<b>100.00</b>