



HAL
open science

A SVM framework for fault detection of the braking system in a high speed train

Jie Liu, Yan-Fu Li, Enrico Zio

► **To cite this version:**

Jie Liu, Yan-Fu Li, Enrico Zio. A SVM framework for fault detection of the braking system in a high speed train. *Mechanical Systems and Signal Processing*, 2017, 87, pp.401 - 409. 10.1016/j.ymssp.2016.10.034 . hal-01408781

HAL Id: hal-01408781

<https://hal-centralesupelec.archives-ouvertes.fr/hal-01408781>

Submitted on 20 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A SVM framework for fault detection of the braking system in a high speed train

Jie Liu¹, Yan-Fu Li¹ and Enrico Zio^{1,2}

¹Chair on System Science and the Energetic Challenge, EDF Foundation, Laboratoire Genie Industriel, CentraleSupélec, Université Paris-Saclay Grande voie des Vignes, 92290 Chatenay-Malabry, France

²Energy Departement, Politecnico di Milano, Milano, Italy

Abstract

In April 2015, the number of operating High Speed Trains (HSTs) in the world has reached 3603. An efficient, effective and very reliable braking system is evidently very critical for trains running at a speed around 300 km/h. Failure of a highly reliable braking system is a rare event and, consequently, informative recorded data on fault conditions are scarce. This renders the fault detection problem a classification problem with highly unbalanced data. In this paper, a Support Vector Machine (SVM) framework, including feature selection, feature vector selection, model construction and decision boundary optimization, is proposed for tackling this problem. Feature vector selection can largely reduce the data size and, thus, the computational burden. The constructed model is a modified version of the least square SVM, in which a higher cost is assigned to the error of classification of faulty conditions than the error of classification of normal conditions. The proposed framework is successfully validated on a number of public unbalanced datasets. Then, it is applied for the fault detection of braking systems in HST: in comparison with several SVM approaches for unbalanced datasets, the proposed framework gives better results.

Key words: high speed train; braking system; support vector machine; feature vector selection; threshold optimization; cost-sensitive models; highly imbalanced data; classification

1. Introduction

In recent years, the speed of the new High Speed Trains (HSTs) has increased to more than 300 km/h (He et al., 2013). According to the report of the International Union of Railways, the number of HSTs under operation in the world has reached 3603 in April 2015 (<http://www.uic.org/highspeed>). Safety and reliability in HST operation has attracted much attention (Bondell et al., 1986; Giboni, 2006; He et al., 2013; Fumeo et al., 2015). The braking system is a key safety-relevant component for guaranteeing the efficient deceleration of the train operating at high speed (Wang and Chiueh, 1998).

There are different types of braking systems, e.g. air brakes, vacuum brakes, electro-pneumatic brakes, electronically controlled pneumatic brakes, etc. (<http://www.railway-technical.com/ep-brakes.shtml>). Figure 1 shows the block diagram of a simple electro-pneumatic braking system. Although highly reliable, faults in braking systems may still occur, e.g. stop cock emergency application valve closed, pneumatically brake stop cock closed on bogie, etc. If a fault is detected, proper mitigation actions can be performed timely to avoid serious damages to the train and the passengers. Thus, fault detection in the braking system is clearly an important issue.

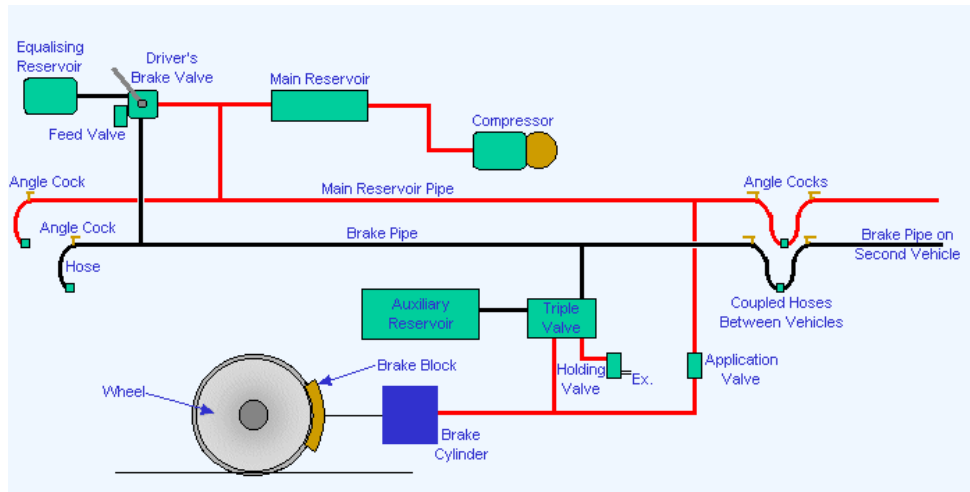


Fig. 1 Block diagram of an electro-pneumatic braking system (<http://www.railway-technical.com/ep-brakes.shtml>).

The increasing installation of monitoring systems on various industrial systems have increased the popularity of data-driven approaches applied for fault detection (Cecati et al., 2015; Gupta et al., 2015; Li et al., 2015). In the braking system of a HST, different variables directly or indirectly related to condition of the braking system are monitored, e.g. battery level in the auxiliary system, line voltage, mean suspension pressure, etc.

In this paper, we propose a framework for fault detection in the braking system of a HST. The framework stands on a modified version of Support Vector Machine (SVM). SVM (Zhao et al., 2015; Zhao et al., 2013) is a powerful data-driven approach that has been already successfully used for fault detection and diagnosis in different applications, e.g. smart grid (Jindal et al., 2016), automobile hydraulic brake system (Jegadeeshwaran and Sugumaran, 2015), gasoline engine valve (Li et al., 2012), bearings (Liu et al., 2016; Lei et al., 2016). SVM solves nonlinear problems by mapping the data into a high-dimensional feature space, i.e. Reproduced Kernel Hilbert Space (RKHS), where the problem becomes linear. The kernel function, which represents the inner product of the mapping of two data points in RKHS, is an important part of a SVM model, as the mapping function is normally difficult to be given explicitly.

The rest of the paper is structured as follows. Research on braking systems and SVM approaches for unbalanced data are reviewed in Section 2. Section 3 presents the proposed framework for fault detection. The application on data from a braking system of a HST are given in Section 4. The proposed framework is also verified and validated on a number of public unbalanced datasets in this section. Conclusions are drawn in Section 5, with some perspectives on research and development.

2. Related work

In this Section, published research work on the braking system of HSTs is reviewed and the originality of our development for fault detection is highlighted. SVM approaches that have been developed for classification of unbalanced data are reviewed, some of which are used as benchmark methods for comparison with the proposed framework.

2.1 The braking system of a HST

HSTs operates at high speed, from 210km/h of the first high speed trains to 350km/h of the new ones (Givoni, 2006). In the literature, there are a few research works on the braking system of a HST, which is a fundamental component for safety. Wang and Chiueh (1998) investigated the brake systems for more efficient operations. Bendell et al. (1986) analyze the reliability of the brake discs on a HST, using a proportional hazard model. An approach combining signal processing and information enhancement is proposed in He et al. (2013) for diagnosis of train bearings. In Kang (2007), a hardware-in-the-loop system is built to simulate the dynamics of a braking system of a HST. A modelling framework of magnetic braking systems is proposed in Galardi et al. (2015) to test efficiency.

To the knowledge of the authors, this is the first time that the fault detection problem of HST braking systems is investigated.

2.2 SVM approaches for unbalanced data

The braking system of a HST is highly reliable. Field data on fault conditions are scarce, so that the fault detection is essentially a classification problem of highly unbalanced data whereby the number of instances in one class (majority class) is much larger than that in another class (minority class). Different SVM approaches have been proposed for tackling unbalanced data. They can be categorized into data-level and algorithm-level ones.

The data-level approaches aim at balancing the data sizes in majority and minority classes. Randomly UnderSampling (RUS) the majority class (Japkowicz 2010) and Randomly OverSampling (ROS) the minority class (Batisata et al., 2004) are two of the most popular data-level approaches. Synthetic Minority Oversampling TEchnique (SMOTE) proposed by Chawla et al. (2002) has been proved to be more effective than ROS.

The algorithm-level approaches modify SVM itself for the classification of unbalanced data. Zhang and Zhou (2015) introduced the Cost-Sensitive SVM (CS-SVM), which applies a relatively larger penalty cost to the error on the minority class and a relatively smaller cost to the error on the majority class, in the objective function of the primal optimization problem of SVM. Yu et al. (2015) attempted to improve the classification accuracy by adjusting the decision boundary, around which different classification decisions are taken.

In this work, we propose a SVM framework in which features and feature vectors are selected by analyzing the between-class separability. In our previous work (Liu and Zio, 2016), a modified version of SVM is proposed with satisfactory results for regression problems. A cost-sensitive version of this method is developed in this paper to tackle the fault detection problem with unbalanced data. A decision boundary optimization method is integrated in the framework to maximize the classification accuracy.

3. SVM framework for fault detection in the braking system of a HST

In this Section, the proposed SVM framework is presented. Figure 2 shows the flowchart of the framework. It can be divided into 4 parts, i.e. feature selection, feature vector selection, model construction and decision boundary optimization.

Suppose the data matrix $\mathbf{x} = \{x_{i,j}\}$ and the vector $\mathbf{y} = \{y_i: y_i \in (-1,1)\}$, for $i = 1, 2, \dots, N_n, N_n + 1, \dots, N_n + N_p, j = 1, 2, \dots, M$ are, respectively, the measured value of the M variables (features) in the braking system and the corresponding fault indicator, with -1 indicating normal conditions and $+1$ indicating faulty condition, and N_n and N_p being the number of data points in normal and faulty conditions, respectively. $x_{i,j}$ is the measured value of the j -th feature in the i -th data point. In the dataset, the first N_n data points are in normal condition and the rest are in faulty condition. In the rest of the paper, the vector $\mathbf{x}_{i:l,j}$ contains the measured value of the j -th feature in the data points from i to l , and the vector $\mathbf{x}_{i,:}$ is the input row vector of the i -th data point.

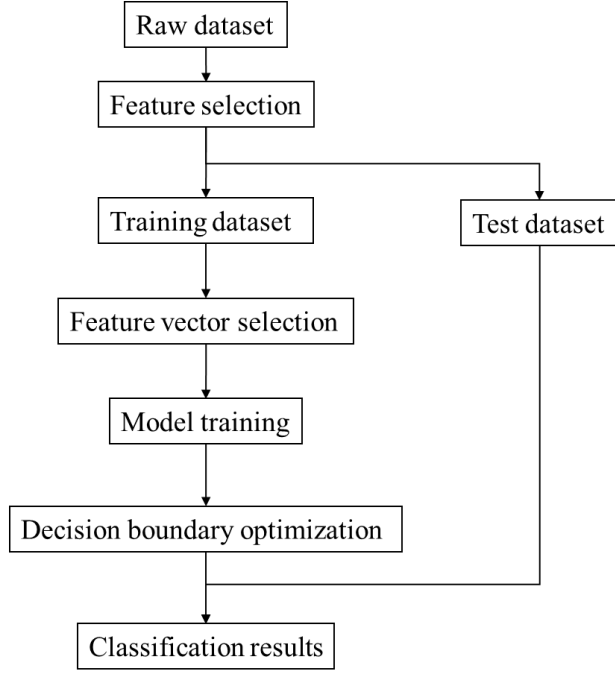


Fig. 2 Flowchart of the proposed SVM framework for fault detection.

3.1 Feature selection

In a dataset containing a large number of features, some features can be redundant or even irrelative to the fault detection. These features need to be identified and eliminated from the data. In the proposed SVM framework, the between-class separability with respect to one feature is used as the criterion. Its mathematical definition with respect to the j -th feature vector is presented as follows:

$$\lambda_j = \frac{(m_{j,+} - m_{j,-})^2}{\delta_{j,+}^2 + \delta_{j,-}^2}, \quad (1)$$

where $m_{j,+}$ and $\delta_{j,+}^2$ are the mean and variance of the vector $\mathbf{x}_{N_n+1:N_n+N_p,j}$, i.e. the values of the j -th feature of the data points in faulty condition (minority class), and $m_{j,-}$ and $\delta_{j,-}^2$ are the mean and variance of the vector $\mathbf{x}_{1:N_n,j}$, i.e. the values of the j -th feature of the data points in normal condition (majority class). The features giving a between-class separability lower than a predefined threshold Th_{fs} are eliminated.

3.2 Feature vector selection

As mentioned in the introduction, SVM solves a nonlinear classification problem by mapping input vectors to a feature vector in RKHS where the problem becomes linear. After selecting the informative features, the data is rearranged by keeping only the selected features. As the data size collected by the monitoring system can be very large, feature vector selection aims at reducing the data size by extracting the informative ones. If one feature vector is to be selected, it should satisfy two criteria:

the new feature vector contains useful information for the classification; the new feature vector brings additional useful information to the current selected feature vectors.

The first criterion is measured by the between-class separability with respect to this feature vector. Suppose that, in SVM, the function for mapping an input vector to a feature vector in RKHS is $(\mathbf{x}_{i,:})$ and the kernel function is noted as $k(\mathbf{x}_{i,:}, \mathbf{x}_{l,:}) = \langle \varphi(\mathbf{x}_{i,:}), \varphi(\mathbf{x}_{l,:}) \rangle$, $i, l = 1, 2, \dots, N_n, N_n + 1, \dots, N_n + N_p$, where $\langle \blacksquare \rangle$ means the inner production. The calculation of the between-class separability is still as in Equation 1, with j meaning the j -th feature vector. But $m_{j,+}$, $\delta_{j,+}^2$, $m_{j,-}$ and $\delta_{j,-}^2$ are calculated as below (Mao, 2002):

$$\begin{aligned} m_{j,+} &= \frac{1}{N_p} \sum_{i=N_n+1}^{N_n+N_p} k(\mathbf{x}_{i,:}, \mathbf{x}_{j,:}) \\ \delta_{j,+}^2 &= \frac{1}{N_n} \sum_{i=N_n+1}^{N_n+N_p} (k(\mathbf{x}_{i,:}, \mathbf{x}_{j,:}) - m_{j,+})^2 \\ m_{j,-} &= \frac{1}{N_n} \sum_{i=1}^{N_n} k(\mathbf{x}_{i,:}, \mathbf{x}_{j,:}) \\ \delta_{j,-}^2 &= \frac{1}{N_n} \sum_{i=1}^{N_n} (k(\mathbf{x}_{i,:}, \mathbf{x}_{j,:}) - m_{j,-})^2 \end{aligned} \quad (2)$$

If the between-class separability of all the data points with respect to the feature vector of one data point is larger than a predefined threshold Th_{fvs} , this data point contains useful information for the classification of the unbalanced data. Otherwise, it is not useful as it fails to distinguish the two classes.

The second criterion is characterized by the variable Local Fitness (LF) proposed by Baudat and Anouar (2003). Suppose that the current selected feature vectors are $\mathbf{s} = \{\varphi(\mathbf{x}_{1,:}), \varphi(\mathbf{x}_{2,:}), \dots, \varphi(\mathbf{x}_{M,:})\}$ and one needs to know if a new data point $\mathbf{x}_{i,:}$ carries additional information to \mathbf{s} ; we attempt to verify if the feature vector $\varphi(\mathbf{x}_{i,:})$ of $\mathbf{x}_{i,:}$ can be expressed as a linear combination of the current feature vectors in \mathbf{s} , as expressed as the below equation:

$$\text{minimize}_{\beta_j} \gamma(\mathbf{x}) = \frac{\|\varphi(\mathbf{x}_{i,:}) - \sum_{j=1}^M \beta_j \varphi(\mathbf{x}_{j,:})\|^2}{\|\varphi(\mathbf{x}_{i,:})\|^2}. \quad (3)$$

The minimum of Equation (3) is given by

$$\boldsymbol{\beta} = K_{\mathbf{S},\mathbf{x}}^t K_{\mathbf{S},\mathbf{S}}^{-1} \mathbf{s}, \quad (4)$$

with $\boldsymbol{\beta} = \{\beta_j\}, i = 1, 2, \dots, M$.

The LF is defined as

$$LF(\mathbf{x}_{i,:}) = \left| 1 - \frac{K_{\mathbf{S},\mathbf{x}_{i,:}}^t K_{\mathbf{S},\mathbf{S}}^{-1} K_{\mathbf{S},\mathbf{x}_{i,:}}}{k(\mathbf{x}_{i,:}, \mathbf{x}_{i,:})} \right|, \quad (5)$$

with $K_{S,S}$ being the kernel matrix of S , \mathbf{K}^t being the transpose of matrix \mathbf{K} and $K_{S,x_{i,:}} = \{k(x_{i,:}, x_{j,:})\}, j = 1, 2, \dots, M$. If $LF(x_{i,:}) > 0$, $\varphi(x_{i,:})$ is a new feature vector, as it can not be expressed as a combination of the current selected feature vectors, i.e. it brings additional information to S ; otherwise, $\varphi(x_{i,:})$ is not a new feature vector, as $\varphi(x_{i,:}) = \sum_{j=1}^M \beta_j \varphi(x_{j,:})$.

The pseudo-code of the feedforward feature vector selection process is shown in Figure 3.

Initialization:

Training dataset: $T = \{(x_{i,:}, y_i)\}$, for $i = 1, 2, \dots, N_n + N_p$

Feature space: $S = []$

Threshold of LF: Th_{lf} (a small positive value)

FVS:

Step 1: First feature vector in S :

For $i = 1$ to $N_n + N_p$

$S = \{x_{i,:}\}$, compute the **separability** of all training data points with respect to the present S , following Equations 1 and 2.

End for.

Select the point which gives the maximal **separability** as the first feature vector and add it to S as the first FV. T is reduced to the complement of E in T i.e. $T = T \setminus E$, with $E = \{(x_{i,:}, y_i) : LF(x_{i,:}) \leq Th_{lf}\}$.

Step 2: Second and following feature vectors:

If T is not empty

Select the data point k from T which gives the maximal **separability** of all the data points as the next feature vector candidate;

If the separability of the all the data points with respect to this data point, i.e. λ_k is bigger than Th_{fvs} , add it to S and reduce the dataset $T = T \setminus E$, with $E = \{(x_{i,:}, y_i) : LF(x_{i,:}) \leq Th_{lf}\}$;

Otherwise, exit the feature vector selection process.

End if

Fig. 3 Pseudo-code of the feature vector selection process.

The selected feature vector should satisfy two requirements: it carries additional information, i.e. $LF(x_{i,:}) > Th_{lf}$; it gives largest separability of all the data points among the ones that satisfy the first requirement, and its separability should be large than the predefined threshold Th_{fvs} .

The threshold for LF, i.e. Th_{lf} is a small positive value to reduce the influence of the noise on the feature vector selection process.

3.3 Model construction

Liu and Zio (2016a) propose a modified version of the least squared SVM (named Feature Vector Regression (FVR)) based on the feature vector selection method proposed in Baudat and Anouar (2003). The estimate function in FVR is a kernel expansion on the selected feature vectors and the objective is to minimize the prediction error on the whole dataset. It exhibits a good generalizability, accuracy and efficiency in the experiments.

In this paper, based on the new feature vector selection method proposed in Section 3.2 for classification, FVR is extended to a cost-sensitive version for the classification of unbalanced data.

Suppose that the feature vectors selected by the method proposed in Section 3.2 are noted as $(\mathbf{x}_{j,:}, \mathbf{y}_j), j = 1, \dots, M$; the primal optimization problem is formulated as

$$\begin{aligned} \text{minimize}_{\hat{y}_j, b} \quad & W = \sum_{i=1}^{N_n} (g(\mathbf{x}_{i,:}) - y_i)^2 + \omega \sum_{i=N_n+1}^{N_n+N_p} (g(\mathbf{x}_{i,:}) - y_i)^2 \\ \text{subject to} \quad & g(\mathbf{x}_{i,:}) = \sum_{j=1}^M \beta_j(\mathbf{x}_{i,:})(\hat{y}_j - b) + b, i = 1, 2, \dots, N_n + N_p \end{aligned} \quad (6)$$

with $\hat{y}_j, j = 1, 2, \dots, M$ being the predicted value of the selected feature vectors, $\beta_j(\mathbf{x}_i), j = 1, \dots, M$ calculated with Equation 4, b a constant value and $\omega > 1$ the cost assigned to the classification error on the positive (minority) class. The estimate function of FVR is $g(\mathbf{x}_{i,:}) = \sum_{j=1}^M \beta_j(\mathbf{x}_{i,:})(\hat{y}_j - b) + b$, with $\hat{y}_j, j = 1, 2, \dots, M$ and b the unknown values.

The unknown values can be calculated analytically by exploiting the Karush–Kuhn–Tucker (KKT) conditions of the dual problem of Equation 6. The dual problem of Equation 6 is

$$W = \sum_{i=1}^{N_n} (\sum_{j=1}^M \beta_j(\mathbf{x}_{i,:})(\hat{y}_j - b) + b - y_i)^2 + \omega \sum_{i=N_n+1}^{N_n+N_p} (\sum_{j=1}^M \beta_j(\mathbf{x}_{i,:})(\hat{y}_j - b) + b - y_i)^2 \quad (7)$$

Supposing $\rho_i = 1 - \sum_{j=1}^M \beta_j(\mathbf{x}_i)$, the KKT conditions of Equation 7 give

$$\begin{aligned} \frac{\partial W}{\partial \hat{y}_{j_0}} &= \sum_{j=1}^M (\sum_{i=1}^{N_n} \beta_{j_0}(\mathbf{x}_i) * \beta_j(\mathbf{x}_i) + \omega \sum_{i=N_n+1}^{N_n+N_p} \beta_{j_0}(\mathbf{x}_i) * \beta_j(\mathbf{x}_i)) * \hat{y}_j + b * (\sum_{i=1}^{N_n} \beta_{j_0}(\mathbf{x}_i) * \rho_i + \\ &\omega \sum_{i=N_n+1}^{N_n+N_p} \beta_{j_0}(\mathbf{x}_i) * \rho_i) - (\sum_{i=1}^{N_n} \beta_{j_0}(\mathbf{x}_i) * y_i + \omega \sum_{i=N_n+1}^{N_n+N_p} \beta_{j_0}(\mathbf{x}_i) * y_i) = 0, \text{ for } j_0 = 1, \dots, M, \quad (8) \\ \frac{\partial W}{\partial b} &= \sum_{j=1}^M (\sum_{i=1}^{N_n} \beta_j(\mathbf{x}_i) * \rho_i + \omega \sum_{i=N_n+1}^{N_n+N_p} \beta_j(\mathbf{x}_i) * \rho_i) * \hat{y}_j + b * (\sum_{i=1}^{N_n} \rho_i^2 + \omega \sum_{i=N_n+1}^{N_n+N_p} \rho_i^2) - \\ &(\sum_{i=1}^{N_n} \rho_i * y_i + \omega \sum_{i=N_n+1}^{N_n+N_p} \rho_i * y_i) = 0. \quad (9) \end{aligned}$$

The unknown values, i.e. $\hat{y}_j, j = 1, 2, \dots, M$ and b , are calculated with

$$\begin{bmatrix} \mathbf{\Omega} & \mathbf{H} \\ \mathbf{\Gamma}^T & c \end{bmatrix} \begin{bmatrix} \hat{\mathbf{y}} \\ b \end{bmatrix} = \begin{bmatrix} \mathbf{P} \\ l \end{bmatrix}, \quad (10)$$

where $\mathbf{\Omega}$ is a $M \times M$ matrix with $\mathbf{\Omega}_{mn} = \sum_{i=1}^{N_n} \beta_m(\mathbf{x}_i) * \beta_n(\mathbf{x}_i) + \omega \sum_{i=N_n+1}^{N_n+N_p} \beta_m(\mathbf{x}_i) * \beta_n(\mathbf{x}_i)$, \mathbf{H} is a $M \times 1$ vector with $\mathbf{H}_m = \sum_{i=1}^{N_n} \beta_m(\mathbf{x}_i) * \rho_i + \omega \sum_{i=N_n+1}^{N_n+N_p} \beta_m(\mathbf{x}_i) * \rho_i$, $\mathbf{\Gamma}$ is a $M \times 1$ vector with $\mathbf{\Gamma}_m = \sum_{i=1}^{N_n} \beta_m(\mathbf{x}_i) * \rho_i + \omega \sum_{i=N_n+1}^{N_n+N_p} \beta_m(\mathbf{x}_i) * \rho_i$, c is a constant and $c = \sum_{i=1}^{N_n} \rho_i^2 + \omega \sum_{i=N_n+1}^{N_n+N_p} \rho_i^2$; $\hat{\mathbf{y}} = (\hat{y}_j), j = 1, 2, \dots, M$ and b are the unknown values in Equation 6, \mathbf{P} is a $M \times 1$ vector with $\mathbf{P}_m = \sum_{i=1}^{N_n} \beta_m(\mathbf{x}_i) * y_i + \omega \sum_{i=N_n+1}^{N_n+N_p} \beta_m(\mathbf{x}_i) * y_i$, $l = \sum_{i=1}^{N_n} \rho_i * y_i + \omega \sum_{i=N_n+1}^{N_n+N_p} \rho_i * y_i$.

3.4 Decision boundary optimization

For the constructed model, the predicted value of a new data point $\mathbf{x}_{i,:}$ is given by $g(\mathbf{x}_{i,:}) = \sum_{j=1}^M \beta_j(\mathbf{x}_{i,:})(\hat{y}_j - b) + b$. In general, a decision boundary τ is predefined and if $g(\mathbf{x}_{i,:}) \geq \tau$, the class label of the new data point is positive (the minority class); otherwise, its class label is negative (the majority class). As in Yu et al. (2015), the decision boundary can be optimized during the training process by maximizing the classification accuracy. With TP, TN, FP, FN representing the number of true positive, true negative, fault positive and fault negative, respectively, two of the most popularly used classification measures are

$$F\text{-measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

and

$$G\text{-mean} = \sqrt{\text{TPR} * \text{TNR}}, \quad (12)$$

where $\text{Precision} = TP / (TP + FP)$, $\text{Recall} = TPR = TP / (TP + FN)$ and $\text{TNR} = TN / (TN + FP)$.

In many papers, G-mean is used as the classification measure for optimizing the decision boundary. G-mean characterizes the classification accuracy on the positive and negative classes. For unbalanced data, especially the highly unbalanced case, a high G-mean normally implies a high false alarm rate, as the number of FP is much larger than that of the TP. In industrial practice, a high fault detection rate and a low false alarm rate are both very important. Thus, in this paper, the sum of F-measure and G-mean is used as the classification accuracy metric to optimize the decision boundary of the constructed model.

4. Application results

In this section, the proposed SVM framework is firstly validated on 15 public datasets from KEEL-dataset repository (Alcala-Fdez et al., 2009) and, then, the results of fault detection of a braking system in a HST are presented.

4.1 Model validation

Fifteen KEEL datasets with different Imbalance Ratio (IR) are used to verify the effectiveness of the model proposed in Sections 3.2, 3.3 and 3.4. The characteristics of the KEEL datasets are shown in Table 1.

Table 1. Characteristics of the KEEL datasets for verifying the effectiveness of the proposed model.

Dataset	Number of instances	Number of attributes	Imbalance Ratio (IR)
glass1	214	9	1.82
haberman	306	3	2.78
new-thyroid1	215	5	5.14
yeast3	1484	8	8.1
ecoli3	336	7	8.6
ecoli-0-6-7_vs_5	220	6	10
yeast-1_vs_7	459	7	14.3
ecoli4	336	7	15.8
abalone-9_vs_18	731	8	16.4
shuttle-6_vs_2-3	230	9	22
yeast4	1484	8	28.1
yeast5	1484	8	32.73
poker-8-9_vs_5	2075	10	82
poker-8_vs_6	1477	10	85.88
abalone19	4174	8	129.44

Comparisons are carried out with three of the most popular SVM approaches for unbalanced data: SVM-RUS, SVM-SMOTE and CS-SVM. Ten-folds cross validation is used to train the models. Grid search is used to tune the SVM hyperparameters in each model. Each dataset is divided into ten equal-size folds, where one fold is selected as test dataset and the rest nine folds form the training dataset. This is repeated ten times and Table 2 reports the mean values of the F-measure and G-mean on the test datasets. The bolded value shows the best results obtained by the four approaches. In the experiment, the proposed framework outperforms the other three benchmark methods for respectively 12 and 9 out of the 15 datasets with respect to F-measure and G-mean. Friedman test and Bonferroni-Dunn test for the experiment results (Liu and Zio, 2016b) show that the proposed framework gives significantly better results than the benchmark methods. One can observe that for the dataset glass1, with respect to F-measure and G-mean the proposed approach gives worse results than all the benchmark methods. This is due to the FVS process, which causes loss of information by reducing drastically the size of the dataset with a small IR (i.e. 1.82 for dataset glass1: the dataset is nearly a balanced one); on the

contrary, the size of the balanced training dataset using RUS, SMOTE is much larger than that using FVS for the dataset glass1.

Table 2. Comparison of the experiment results.

	F-measure				G-mean			
	proposed approach	SVM-RUS	SVM-SMOTE	CS-SVM	proposed approach	SVM-RUS	SVM-SMOTE	CS-SVM
glass1	0.590	0.637	0.641	0.623	0.664	0.714	0.716	0.706
haberman	0.428	0.424	0.405	0.423	0.586	0.533	0.568	0.525
new-thyroid1	0.958	0.881	0.910	0.912	0.980	0.958	0.957	0.955
yeast3	0.773	0.633	0.702	0.668	0.898	0.841	0.866	0.870
ecoli3	0.600	0.522	0.598	0.564	0.760	0.840	0.780	0.793
ecoli-0-6-7_vs_5	0.825	0.495	0.648	0.723	0.859	0.830	0.761	0.808
yeast-1_vs_7	0.391	0.189	0.231	0.291	0.717	0.631	0.539	0.624
ecoli4	0.765	0.594	0.750	0.705	0.884	0.916	0.807	0.790
abalone-9_vs_18	0.667	0.251	0.343	0.349	0.864	0.732	0.593	0.597
shuttle-6_vs_2-3	1.000	0.558	0.877	0.820	1.000	0.751	0.887	0.831
yeast4	0.370	0.196	0.316	0.352	0.727	0.776	0.683	0.678
yeast5	0.635	0.466	0.691	0.695	0.904	0.943	0.881	0.877
poker-8-9_vs_5	0.118	0.035	0.092	0.079	0.669	0.219	0.595	0.511
poker-8_vs_6	0.530	0.251	0.646	0.451	0.619	0.425	0.669	0.489
abalone19	0.071	0.031	0.039	0.046	0.644	0.644	0.346	0.253

4.2 Fault detection of the braking system in a HST

There are totally 43 monitored variables related to the braking system faults. They include train-level conditions, e.g. GPS position, speed, operation mode, external power supply, operation hours, line voltage, line current, and braking system-level conditions, e.g. internal temperature, battery voltage, detected slip or slide, ED brake state, TCL brake state, achieved brake effort. For confidentiality reasons, they are not explicitly listed and are named as Var1, Var2, ..., Var43. As kernel methods can only tackle numeric data, the nominal variables are firstly transformed to numeric ones. All the numeric variables are eventually normalized to [0.1 0.9].

In the braking system of a HST, different faults may occur, e.g. stop cock emergency application valve closed, pneumatically brake stop cock closed on bogie, MTB isolated, stop cock closed, etc. In this paper, the different faults are not distinguished (diagnosed) and the label marks only if the brake system is undergoing a fault (+1) or not (-1).

Not all the monitored variables are useful for the fault detection task. Separability, as introduced in Section 3.1, is used to verify the usefulness of each variable in differentiating the faulty and normal conditions of the braking system, as shown in Table 3.

Table 3. Separability of the faulty and normal conditions with respect to each variable.

Variable name	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9
Separability	0.0749	0.0948	0.0306	0.0023	0.0191	0.0332	0.0030	0.0733	0.0134
Variable name	Var10	Var11	Var12	Var13	Var14	Var15	Var16	Var17	Var18
Separability	0.0111	0.0110	0.0216	0.0073	0.0120	2.4e-6	0.0128	0.0642	0.0079
Variable name	Var19	Var20	Var21	Var22	Var23	Var24	Var25	Var26	Var27
Separability	0.0072	0.0078	0.0002	0.0003	0.0083	0.0995	0.9775	0.0101	0.0099
Variable name	Var28	Var29	Var30	Var31	Var32	Var33	Var34	Var35	Var36
Separability	0.0096	0.0012	0.0135	0.0209	0.0225	0.0254	0.0284	0.0677	0.0349
Variable name	Var37	Var38	Var39	Var40	Var41	Var42	Var43	-----	-----
Separability	0.2768	0.082	0.1175	0.0000	0.0000	0.0000	0.0000	-----	-----

From Table 3, one observes that separabilities of the faulty and normal conditions with respect to Var40-43 are zero. The threshold Th_{fs} for feature selection is set to be 0.1 and, thus, only three variables are selected, i.e. Var25, Var37 and Var39. There are totally 28837 data points in normal conditions and only 159 in faulty conditions. The IR is 181.4, i.e. the data is highly unbalanced.

The data points for normal and faulty conditions with the selected variables are plotted in Figure 3. It is shown that there is a large overlapping of the two conditions and it is difficult to have both high F-measure and G-mean values, i.e. both high precision and recall values, simultaneously. For a relatively high recall, the FP number is high, thus, precision is low, whereas for a relatively high precision, TP is small, i.e. the recall value is low. Simply maximizing the G-mean as in many reported work gives a high recall value but a low precision, i.e. a high false alarm rate, and this is unacceptable for practitioners. In order to have a compromised results, the sum of F-measure and G-mean is used as the classification metric for optimizing the decision boundary, as explained in Section 3.4.

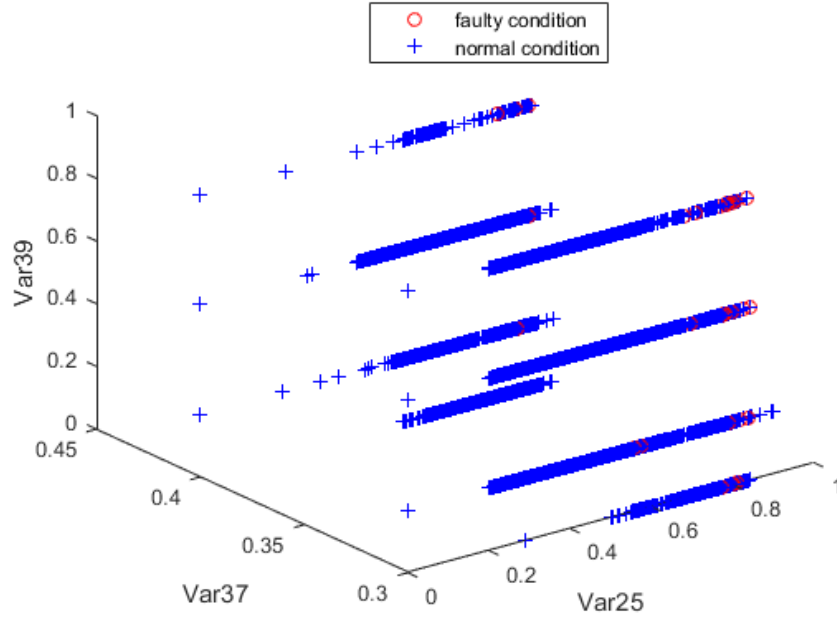


Fig. 3 plot of the normal and faulty data points with respect to the selected features.

Ten-fold cross validation is applied on the dataset. Radial basis kernel function, i.e. $k(\mathbf{x}_{i,:}, \mathbf{x}_{l,:}) = e^{-\|\mathbf{x}_{i,:} - \mathbf{x}_{l,:}\|^2 / \sigma^2}$ is used in the proposed framework. The thresholds, i.e. Th_{fvs} and Th_{lf} for feature vector selection and the parameter in the kernel function σ are tuned by grid search method (Liu et al., 2013). The best results are given by the combination $Th_{fvs} = 0.5$, $Th_{lf} = 10^{-8}$ and $\sigma = 0.9$. All the selected feature vectors give a separability between the normal and faulty conditions bigger than $Th_{fvs} = 0.5$. In comparison with Table 3, one can see that the proposed SVM framework increases the between-class separability than the original features. At the end, only 67 feature vectors are selected, and, thus, the computation is much simplified in comparison with SVM using all the data points.

The classification results with respect to F-measure and G-mean are shown in Table 4. It shows that the proposed framework give a much higher F-measure value and a little lower G-mean value in comparison with the benchmark methods.

Table 4. Classification results of the fault detection system in HST with respect to F-measure and G-mean.

Proposed framework		SVM-RUS		SVM-SMOTE		CS-SVM	
F-measure	G-mean	F-measure	G-mean	F-measure	G-mean	F-measure	G-mean
0.2711	0.8561	0.0457	0.8874	0.0876	0.9092	0.0901	0.9109

For comparison, the precision and recall values obtained by different methods are listed in Table 5. Two values for precision and recall are listed for the proposed framework, because if the decision

boundary is not calculated as the optimized one during the training process, as in Section 3.4, by varying the decision boundary, different pairs of precision and recall values can be obtained (as shown in Figure 4). With the same recall values (0.9874 or 0.9245), the proposed method can give a larger value for the precision, i.e. low false alarm rate. The result given by the proposed framework is more acceptable by the practitioners, as the false alarm rate is relatively low.

Table 5. Classification results of the fault detection system in HST with respect to precision and recall values.

Proposed framework		SVM-RUS		SVM-SMOTE		CS-SVM	
Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
0.0496	0.9874	0.0234	0.9874	0.0460	0.9245	0.0473	0.9245
0.0604	0.9245						

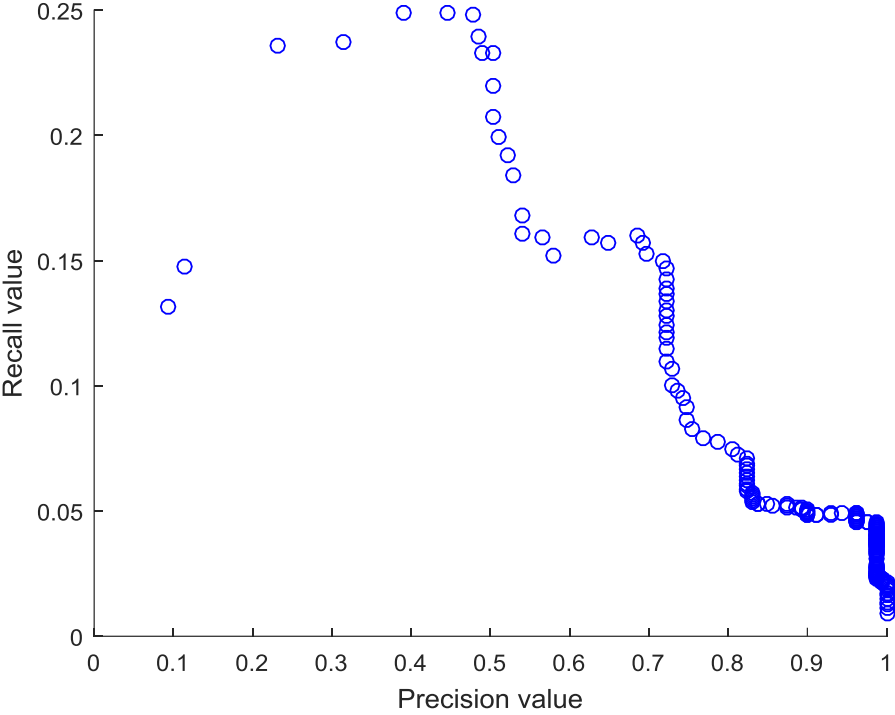


Fig. 4 Precision and recall values for the training dataset with respect to different decision boundaries.

If the decision boundary optimization is optimized according to multiple objectives, with the different pairs of recall and precision values, the decision maker can choose the one that is acceptable for him/her.

5. Conclusions and perspectives

The HST is experiencing a high development period, especially in Asia. The braking system is a critical element for safety and, for this reason, it is highly reliable but may still undergo different types of faults. An efficient fault detection system can be very useful for operators and maintainers. Given the high reliability of the braking system, monitored data on fault conditions is far less than on normal conditions. Then, fault detection becomes a classification problem of highly unbalanced data.

A SVM framework is proposed in this paper for fault detection. The proposed framework includes feature selection, feature vector selection, model construction and decision boundary. It is validated on 15 public datasets, in comparison with three of the most popular SVM models for unbalanced data. The proposed framework outperforms the three benchmark methods for most of the datasets considering the F-measure and G-mean as the performance metrics. The results of application for the fault detection of the braking system show that the proposed method can give a higher precision value than the three benchmark methods, with the same recall value.

Because of the overlapping of the monitored values under normal conditions and faulty conditions, high F-measure and G-mean values can not be achieved simultaneously. If the decision boundary with multi-objective optimization is integrated in the framework, by varying the decision boundary, different choices of precision and recall values (F-measure and G-mean values) can be available for the decision maker to choose, as shown in Figure 4.

The accuracy metric used in this paper is the sum of F-measure and G-mean. Other metrics are available, e.g. Area Under the Curve (AUC). Future work will focus on the investigation of other metrics for characterizing the classification of unbalanced data.

References

- Alcala-Fdez, J., Sanchez, L., Garcia, S., Del Jesus, M.J., Ventura, S., Garrell, J.M., Otero, J., Romero, C., Bacardit, J., Rivas, V.M. and Fernandez, J.C., 2009. KEEL: a software tool to assess evolutionary algorithms for data mining problems. *Soft Computing*, 13(3), pp.307-318.
- Batista, G.E., Prati, R.C. and Monard, M.C., 2004. A study of the behavior of several methods for balancing machine learning training data. *ACM Sigkdd Explorations Newsletter*, 6(1), pp.20-29.
- Baudat, G. and Anouar, F., 2003. Feature vector selection and projection using kernels. *Neurocomputing*, 55(1), pp.21-38.
- Bendell, A., Walley, M., Wightman, D.W. and Wood, L.M., 1986. Proportional hazards modelling in reliability analysis—An application to brake discs on high speed trains. *Quality and reliability engineering international*, 2(1), pp.45-52.

- Cecati, C., Kolbusz, J., Różycki, P., Siano, P. and Wilamowski, B.M., 2015. A novel RBF training algorithm for short-term electric load forecasting and comparative studies. *IEEE Transactions on Industrial Electronics*, 62(10), pp.6519-6529.
- Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, pp.321-357.
- Fumeo, E., Oneto, L. and Anguita, D., 2015. Condition Based Maintenance in Railway Transportation Systems Based on Big Data Streaming Analysis. *Procedia Computer Science*, 53, pp.437-446.
- Galardi, E., Meli, E., Nocciolini, D., Pugi, L. and Rindi, A., 2015. Development of efficient models of Magnetic Braking Systems of railway vehicles. *International Journal of Rail Transportation*, 3(2), pp.97-118.
- Givoni, M., 2006. Development and impact of the modern High-speed train: A review. *Transport reviews*, 26(5), pp.593-611.
- Gupta, S., Kambli, R., Wagh, S. and Kazi, F., 2015. Support-vector-machine-based proactive cascade prediction in smart grid using probabilistic framework. *IEEE Transactions on Industrial Electronics*, 62(4), pp.2478-2486.
- He, Q., Wang, J., Hu, F. and Kong, F., 2013. Wayside acoustic diagnosis of defective train bearings based on signal resampling and information enhancement. *Journal of Sound and Vibration*, 332(21), pp.5635-5649.
- Japkowicz, N., 2000, June. The class imbalance problem: Significance and strategies. In *Proc. of the Int'l Conf. on Artificial Intelligence*.
- Jegadeeshwaran, R. and Sugumaran, V., 2015. Fault diagnosis of automobile hydraulic brake system using statistical features and support vector machines. *Mechanical Systems and Signal Processing*, 52, pp.436-446.
- Jindal, A., Dua, A., Kaur, K., Singh, M., Kumar, N. and Mishra, S., 2016. Decision Tree and SVM-Based Data Analytics for Theft Detection in Smart Grid. *IEEE Transactions on Industrial Informatics*, 12(3), pp.1005-1016.
- Kang, C.G., 2007. Analysis of the braking system of the Korean high-speed train using real-time simulations. *Journal of Mechanical science and Technology*, 21(7), pp.1048-1057.
- Lei, Y., Jia, F., Lin, J., Xing, S. and Ding, S.X., 2016. An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data. *IEEE Transactions on Industrial Electronics*, 63(5), pp.3137-3147.
- Li, N., Zhou, R., Hu, Q. and Liu, X., 2012. Mechanical fault diagnosis based on redundant second generation wavelet packet transform, neighborhood rough set and support vector machine. *Mechanical systems and signal processing*, 28, pp.608-621.
- Li, Z., Outbib, R., Giurgea, S. and Hissel, D., 2015. Diagnosis for PEMFC systems: a data-driven approach with the capabilities of online adaptation and novel fault detection. *IEEE Transactions on Industrial Electronics*, 62(8), pp.5164-5174.
- Liu, R., Yang, B., Zhang, X., Wang, S. and Chen, X., 2016. Time-frequency atoms-driven support vector machine method for bearings incipient fault diagnosis. *Mechanical Systems and Signal Processing*, 75, pp.345-370.
- Liu, J., Seraoui, R., Vitelli, V. and Zio, E., 2013. Nuclear power plant components condition monitoring by probabilistic support vector machine. *Annals of Nuclear Energy*, 56, pp.23-33.

- Liu, J. and Zio, E., 2016. Feature Vector Regression with Efficient Hyperparameters Tuning and Geometric Interpretation. *Neurocomputing*, under submission.
- Liu, J. and Zio, E., 2016. An adaptive online learning approach for Support Vector Regression: Online-SVR-FID. *Mechanical Systems and Signal Processing*, 76, pp.796-809.
- Mao, K.Z., 2002. RBF neural network center selection based on Fisher ratio class separability measure. *IEEE Transactions on Neural Networks*, 13(5), pp.1211-1217.
- Wang, P.J. and Chiueh, S.J., 1998. Analysis of eddy-current brakes for high speed railway. *IEEE Transactions on magnetics*, 34(4), pp.1237-1239.
- Yu, H., Mu, C., Sun, C., Yang, W., Yang, X. and Zuo, X., 2015. Support vector machine-based optimized decision threshold adjustment strategy for classifying imbalanced data. *Knowledge-Based Systems*, 76, pp.67-78.
- Zhang, Y. and Zhou, Z.H., 2010. Cost-sensitive face recognition. *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on*, 32(10), pp.1758-1769.
- Wei, Z., Tao, T., ZhuoShu, D. and Zio, E., 2013. A dynamic particle filter-support vector regression method for reliability prediction. *Reliability Engineering & System Safety*, 119, pp.109-116.
- Zhao, W., Tao, T. and Zio, E., 2015. System reliability prediction by support vector regression with analytic selection and genetic algorithm parameters selection. *Applied Soft Computing*, 30, pp.792-802.