Arc Detection in Pantograph Catenary Systems by the Use of Support Vector Machines Based Classification

Sami Barmada[1], Marco Raugi[1], Mauro Tucci[1],

Francesco Romano[2]

[1]DESTEC, University of Pisa, Italy

[2]Trenitalia s.p.a., Italy

Abstract: Predictive Maintenance, Prognostics and Reliability Centered Maintenance approach are becoming more and more important in the railway sector to reduce costs of operation and increase reliability and safety. In fact they are fundamental to optimize the maintenance process, defining new measures and algorithms to locate faults, monitor health conditions of subsystems and estimate residual life of components.

However there is a tradeoff to identify: new measures and new algorithms imply new sensors and new processing devices, and these items have their own cost and their own reliability; this issue has to be taken into account to evaluate the global benefit.

In some cases, however, it's possible to use existing sensors and existing processing hardware to extract new information from already available data. It's clear that this is usually the best option because the benefit can be achieved with little or not cost at all. This paper describes the result of a study performed with the aim of detecting arcing events without the need of additional equipment mounted on board the train. A set of data relative to voltage and current collected by trains in high-speed lines together with a set of measurements coming from photosensors are available. The data are processed by the use of an advanced classification technique, namely Support Vector Machines, with the aim of extracting important information such as the time coordinate related to anomalies in the overhead contact line and the status of the contact strip of the pantograph.

1. Introduction

Modern high speed trains need a great amount of power and the condition of current collection are critical. Furthermore a severe failure in the current collection system can provoke great damage on the line and serious traffic disruptions. This requires a great care to ensure the efficiency of the contact strips, and implies a continuous cost due to their replacement. In many cases the contact strips are replaced well before their maximum wear limit to avoid risks because there is no measure to evaluate the residual life and their operating conditions. The analysis of the pantograph/catenary subsystem has thereforebeen a hot topic for several years, and is still the object of great interest in the scientific community; furthermore its breakdown would imply a high impact in the economy management of a railway system, and the planned maintenance (usually performed by periodic inspections along the line by the use of an inspection vehicle equipped for catenary checking) is highly expensive. In [1] the importance of condition monitoring and predictive analysis is well described and a set of techniques for the monitoring of different components of the railway infrastructure are described.

High quality current collection is characterized by a continuous contact between the pantograph and the overhead line while poor contact behavior produces arcing, wear. etc.

Some approaches analyze the mechanical behavior (i.e. vibrations) of the catenary, measured by experimental setup located either on a train or in a fixed place over the contact line [2] - [5]. This approach is very accurate when macroscopic defects are present while not always small defects are evidenced. In [6] a very interesting system is proposed for the condition monitoring of the pantograph – catenary system, still based on the presence of accelerometers, to be mounted directly on a train and integrated in a general data acquisition platform.

Other experimental techniques have been proposed, in which the train is equipped with either a phototube or a photodiode capable of detecting electric arcing through the ultraviolet emission. In general a couple of twin phototube sensors are physically placed on the top deck of the locomotive at the front and rear part of the pantograph. The results are positive but this

technique is not convenient for a large scale application because of economical reasons [7] – [10].

Recently some attempts to analyze the collected current have been made; in particular, the Fourier Transform (FT) is used for the analysis of the current. The method explained in [11] gives only partial results, since the FT hides the information regarding time (i.e. the information regarding the location of the defect) and is not always effective, as it will be shown later.

A noteworthy contribution to the research has been given in the past by the authors for a dc railway system. The data obtained by a phototube are compared to the current collected by the train; an advanced data analysis technique based on wavelet expansion, makes it possible to detect and locate with high accuracy the presence of the welding effect by simply analyzing the collected current [12]. This is clearly a great improvement on previous techniques since it avoids the need of a photosensitive device.

The authors have applied the same wavelet based technique and the Hilbert-Huang transform to the AC signals in [13], but the results have not beenas promising as in the DC case, probably due to the higher complexity of the arcing phenomena in AC (i.e. the arc does not always ignite at the same voltage and a nonlinear behavior can appear). For this reason the authors have investigated, in this work, other approaches, namely Support Vector Machines (SVM). SVM are capable of classifying events when they can be trained on a statistically significant data set, and can be useful for such analysis since no physical knowledge of the phenomenon is requested, which is important in an industrial environment where the only important aspect is the efficiency of the preventive maintenance. The main goal is to train the SVM by using voltage and/or current data and phototube data; in this way the SVM (once trained) should be able to detect the presence of an electric arc by only working on voltage and/or current, without the need of the phototube. In this way, by processing data regularly available for high speed trains, it would be possible to obtain useful information on the pantograph/catenary state without the need of additional equipment or inspections.

It is worth to note that at the author's knowledge, little or no previous works may be found in the literature reporting the use of supervised classification techniques to detect arcs in pantograph catenary systems directly from the currents and voltages measurements readily available on the train. In this regard, this work represents a first research step in this direction.

In section II the basic principles of SVM classifications are outlined, then in section III the description of the results obtained by the analysis is given. In section IV a comparison between the proposed SVM approach and RBF networks is shown.

2. Classification of time domain signals

2.1 Data pre-processing

Classification algorithms of time series in general do not use the raw data signals in the time domain, but require a reduced dimensional input vector that represents each time series. To obtain an input dataset, a feature extraction has been performed on the time domain signals, which is based on the calculation of the periodogram. Given a generic time series $x(k)$, at times $k = 1..m$, the periodogram $X(q)$ at frequencies $q = 1..m$ is defined as

$$X(q) = \frac{1}{m}\left|\sum_{k=1}^{m} x(k) e^{-j2\pi qk\frac{1}{m}}\right|^2 \tag{1}$$

Periodogram analysis gives good results in supervised and unsupervised classification of time series data [10], and it is also very easy to compute using an FFT algorithm. In general a reduced number of samples of the periodogram gives a good representation of the time series, in particular a metric based on truncated periodogram in logarithmic scale is proposed in [14].

The time position k_i of a number of arc events $i = 1\ldots N$ is detected by using the photosensor signal, and for each occurrence we compute the periodogram of the *2m* samples time window $[k_i - m, k_i + m - 1]$ of the voltage or current signal. These frequency domain signals are used to identify the arc events. A number of N periodogram signals that are not correlated with arc events are also computed to represent the signal in the case that the arc is not present. The total number of input data signals is then $2N$.

The logarithmic periodogram signals $\log_{10}(X(q))$ are truncated to their first p components, to form the input vectors $\mathbf{x}_i \in \Re^p$, $i = 1..2N$, while the target vector components $\mathbf{y} \in \Re^{2N}$ indicates if the input $\mathbf{x}_i$ is related to an arc event or not, correspondingly $y_i = \{1, -1\}$.

2.2 Support Vector Machine based classification

The Support Vector Machine is a supervised classification technique [15]. The formulation of the SVM classifier used in this work is the soft margin nonlinear classification [15],[16], which is based on the following optimization problem (primal problem):

$$\begin{aligned} &\min_{\mathbf{w},b,\xi} \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{2N}\xi_i \\ &\text{subject to } y_i\left(\mathbf{w}^T\phi(\mathbf{x}_i) + b\right) \geq 1 - \xi_i \\ &\xi_i \geq 0,\ i = 1\ldots 2N \end{aligned} \tag{2}$$

where the nonlinear transform $\phi(\mathbf{x}_i)$ maps $\mathbf{x}_i$ into a higher dimensional feature space, and $C \geq 0$ is the regularization parameter. The vector $\mathbf{w}$ denotes the normal vector to the optimal separation hyperplane in the transformed space, whereas slack variables ξ_i measure the degree of misclassification of the vectors $\mathbf{x}_i$. The decision function is then given by

$$\operatorname{sgn}\left(\mathbf{w}^T\phi(\mathbf{x}) + b\right) \tag{3}$$

The solution of the constrained problem in (2) is obtained by the method of Lagrange multipliers, in particular by solving the dual problem which is in the form:

$$\begin{aligned} &\min_{\alpha} \frac{1}{2}\alpha^T Q\alpha - \sum_{i=1}^{2N}\alpha_i \\ &\text{subject to } \sum_{i=1}^{2N} y_i\alpha_i = 0 \\ &0 \leq \alpha_i \leq C, i = 1\ldots 2N \end{aligned} \tag{4}$$

where $\alpha \in \Re^{2N}$ is the vector of the Lagrange multipliers, and Q is a positive semidefinite $2N$ by $2N$ matrix $Q_{i,j} = y_i y_j k(\mathbf{x}_i\mathbf{x}_j)$, where k is the kernel function that represents the dot product in the transformed high dimensional space.

The advantages of using the dual form are that the calculation of the transform $\phi(\mathbf{x}_i)$ is not needed, as only the calculation of kernel function is required (kernel trick), and that the slack variables ξ_i vanish, with the constant C appearing only as an additional constraint on the Lagrange multipliers.

After problem (4) is solved, only a few α_i will be greater than zero, and the corresponding $\mathbf{x}_i$ are the support vectors. Using the primal - dual relationships theoptimal $\mathbf{w}$ is obtained as:

$$\mathbf{w} = \sum_{i=1}^{2N} y_i \alpha_i \phi(\mathbf{x}_i) \tag{5}$$

The bias b is calculated as

$$b = \frac{1}{N_{SV}} \sum_{i \in SV} \mathbf{w}^T \phi(\mathbf{x}_i) - y_i \tag{6}$$

where $SV \equiv \{i : \alpha_i > 0\}$ is the set of the support vectors indices, and N_{SV} is the number of support vectors.

The dot product $\mathbf{w}^T \phi(\mathbf{x}_j)$ needed for calculating the bias and also decision function (3) is calculated using the kernel trick

$$\mathbf{w}^T \phi(\mathbf{x}_j) = \sum_{i=1}^{2N} y_i \alpha_i \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) = \sum_{i=1}^{2N} y_i \alpha_i k(\mathbf{x}_i, \mathbf{x}_j) \tag{7}$$

In this way the direct calculation of $\phi(\mathbf{x}_i)$ is never required and the classification function (3) becomes

$$\operatorname{sgn}\left(\sum_{i=1}^{2N} y_i \alpha_i k(\mathbf{x}_i, \mathbf{x}_j) + b \right) \tag{8}$$

The solution of the problem (4) is performed by means of Sequential Minimal Optimization (SMO) [17].

The SVM method presented requires the selection of the kernel, and the parameter C. In this work we use the Gaussian kernel, or radial basis function RBF,

$$k_{RBF}\left(\mathbf{x}_i,\mathbf{x}_j\right)=\exp\left(-\gamma\left\|\mathbf{x}_i-\mathbf{x}_j\right\|^2\right) \tag{9}$$

which requires the choice of the parameter γ.

3. Data Analysis

3.1 Data description and pre processing

The data available for the analysis are relative to 7 test runs of 25 kV a.c. high speed trains, operated on regular passengers railway tracks. The trains are equipped with voltage and current recording instruments (which are always present on high speed trains), and two phototubes revealing the presence of the electric arcs. In particular, for one out of seven test runs (named run #1) voltage and current are sampled at 5 kHz, while for the other 6 runs (named runs #2 to #7) they are sampled at 20 kHz. The authors will take into account this difference in the results section, since it would be important that the proposed technique can be implemented economically, hence a lower sampling rate for voltage and current would mean cheaper and less complex recording instruments.

In particular, for each test run the data available are:

- Voltage.
- Two currents (due to the fact that there are two parallel circuits feeding the 4 electrical engines).
- Two phototubes (looking at the same pantograph from both directions).

In each test run the above described quantities have been recorded for approximately 25 minutes.

In Figure 1 the signal of the phototube in presence of an arc is shown, while in Figures 2 and 3 the corresponding voltage and current are shown. It is evident, comparing Figures 1 to 3, that a direct analysis of current or voltage would not lead to any significant conclusion.

Figure 4 shows the logarithmic periodogram of four current signals where p=11 and m=50; Two signals correspond to strong arc events and two signals correspond to time intervals where no arc effect is present in the phototube output. It can be observed from

Figure 4 that cases 1 and 2 are very similar, yet they are related to opposite events while case 3 presents significant differences to the other three cases. In general it is not trivial to relate a frequency pattern to the presence of an arc, this is why a straightforward use of the Fourier Transform does not often gives the desired answers for this particular problem. Therefore a classification procedure is necessary and would give potentially better results.

Commonly, electric arcs whose durations are shorter than 5*ms* (according to the phototube measurement) are neglected when the data are used for the catenary status assessment. In this proposal we do not make this distinction, since the presence of shorter duration events might indicate performance degradation of the pantograph contact strip due to excessive wear; therefore each arc collected by a phototube is considered an important event to be identified and classified.

Consequently a variable threshold is used to identify the arcs and the threshold is obtained by means of biasing a moving average of the photosensor signal, as shown in Figure 5. Arcs (and its time coordinate) are identified when the signal of one of the two photosensors exceeds the threshold.

3.2 Determining the most informative input signals

In the first case study the photosensor, currents and voltage signals are sampled at 5 kHz and a total number of around $7 \bullet 10^6$ points are recorded for each signal; Figure 6 shows the complete signal relative to one phototube. As a first step the photosensors signals are analyzed in order to identify the arc events using the variable threshold method described above, where the moving average window is set to 6000 points, and the bias is selected in order to avoid the background noise. As a result a total number of $N = 22039$ arcs are identified using both phototubes, hence only the 0.3% of the recorded signal is affected by the presence of electric arcs. For each identified arc $i = 1 \ldots N$ we save its time position index k_i; in addition, a number N of time positions k_i for $i = N+1 \ldots 2N$ where the arc is not present are randomly selected from the signal. Therefore the SVM for classification will be trained

using $2N = 44078$ points, with two balanced classes.

The scope of the classification is to detect the arc presence from the currents and voltage signals: once the SVM is trained with known data, it is able to assign an input time series (voltage and/or current) to one of the two classes (arc present or not present) without the need of the photosensor.

Due to the big amount of data available from direct measurements (one voltage signal V_1 and two currents signals I_1, I_2) we need to determine which signals are more correlated to the arc occurrences; therefore as a preliminary analysis we train a SVM using different combinations of the input signals, in order to determine which combination gives higher accuracy results.

As described in section 2.1 we associate each occurrence $i = 1 \ldots 2N$ to the correspondent time series of $2m$ samples of the currents and voltage signals around the time positions k_i. For each time series we calculate its logarithmic periodogram, and we truncate it to the first p components in order to obtain the SVM inputs. For this preliminary analysis we use $m = 50$ and $p = 6$. Note that with this choice each time series has a length of $20\,ms$, which corresponds to one period of the main signal. In this analysis we use base parameters for the SVM, with a Gaussian kernel, in particular $C = 1, \gamma = \frac{1}{p}$.

To evaluate the accuracy of a SVM classification we use the n-fold cross validation (CV), which is defined as the ratio between the number of correctly identified points and the total number of points in test data, and averaging over n-folds.

Table I shows the results of the 3-fold cross-validation using different input signals. From row #1 to row #4 the input vector is the truncated logarithmic periodogram of $2m$ samples of the indicated quantity; the sum of the two currents $I_1 + I_2$ is the signal which gives more accurate results. Rows #5 to row #8 show the results of the analysis performed by concatenating two different signals; in this case the input vector is the concatenation of the

logarithmic truncated periodograms of the signals, and in this case $\gamma = \frac{1}{2p}$. The use of the combination of voltage and the sum of the currents gives the best result in our analysis (more than 86% of the events are correctly evidenced).

In the last column we use the concatenation of the three signals (voltage and the two currents) obtaining the higher dimension of input vector (3*p*) leading to a value of $\gamma = \frac{1}{3p}$.

Table I

Preliminary analysis

Signal	**3-fold CV**
V_1	79.4236%
I_1	79.7276%
I_2	78.8723%
$I_1 + I_2$	80.3061%
$[V_1, I_1]$	85.3925%
$[V_1, I_2]$	86.0368%
$[V_1, I_1 + I_2]$	86.2614%
$[I_1, I_2]$	83.6865%
$[V_1, I_1, I_2]$	84.3671%

3.3 Analysis of test run #1

Following the previous results we search for the best SVM parameters using the best input configuration $[V_1, I_1 + I_2]$.

As performance index we use now a 5-fold CV, and we perform a grid-search approach [18] to obtain the best (C, γ) pattern for different choices of the following parameters:

- temporal window *m* = 50, 100,150,200
- logarithmic periodogram truncation (frequency window): *p* = 6, 12

- kernel type: RBF, as in (9) or sigmoidal, namely $k_{sig}\left(\mathbf{x}_i,\mathbf{x}_j\right)=\tanh\left(\gamma\,\mathbf{x}_i^T\cdot\mathbf{x}_j\right)$

Table II

grid search for best accuracy

Kernel	**Frequency window *p***	**Time window *m***	**5-fold CV**
RBF	6	50	87.3025
RBF	6	100	90.0481
RBF	6	150	90.8196
RBF	6	200	88.4597
RBF	12	50	87.6882
RBF	12	100	90.5473
RBF	12	150	92.6576
RBF	12	200	88.9589
SIG	6	50	83.7400
SIG	6	100	88.2328
SIG	6	150	89.3220
SIG	6	200	85.3283
SIG	12	50	85.8956
SIG	12	100	89.7304
SIG	12	150	90.5019
SIG	12	200	85.8049

Various approaches have been proposed in literature to determine the SVM parameters, and some analytical procedures exist in particular for SVM regression methods [19]. For classification problems the selection of the parameter pair (C,γ) is a hard task, which has been approached also using evolutionary optimization methods [20]. One of the most robust, efficient and well accepted techniques is the two step grid search approach described in [18], which consists in calculating the CV accuracy first in a coarse grid of exponentially growing values of (C,γ), and then, a second grid search is performed on a finer grid in the region where the better results have been found in the first step. In the following of this work all the reported SVM results are obtained using the two step grid search approach, which for instance can be easily parallelized.

Table II shows the 5-fold CV obtained in the different cases described above; we can observe that the RBF kernel in general outperforms the sigmoidal kernel, and that the 5-fold CV increases for increasing m from 50 to 150 but it decreases at 200, hence the best time window is of $2m=300$ points, corresponding to three mains periods. Increasing the frequency window p from 6 to 12 gives in general a small improvement of the 5-fold CV. The

best configuration is highlighted in Table II. For this choice the input of the SVM has dimension $2p = 24$, and it is worth to note that in general the SVM algorithm is not significantly affected by the course of dimensionality, with respect to the input dimension. So choosing $p = 12$ gives slightly better accuracy without significantly increasing the computational costs with respect to the smaller frequency window. For the best configuration highlighted in the table the SVM parameters found by using grid search are the following : $C = 4871$, $\gamma = 8.6317 \cdot 10^{-5}$. It is important to note that, with a sampling frequency of 5kHz and a window of 300 points, the frequency resolution of the periodogram is 16.67 Hz . So the third harmonic is the mains period of 50Hz, and with 12 points of the periodogram we look up to 183.3 Hz.

The obtained accuracy of around 92% means that the SVM, once properly trained with the voltage and current input and phototube results, is capable of correctly classifying the 92% of time windows, detecting the presence or absence of an arc from the analysis of voltage and currents. The remaining 8% is related to false positives and false negatives. This percentage is an extremely good result from an industrial point of view, making the proposed method a potentially powerful instrument.

3.4 Analysis of test runs #2 to #7

In this section other six datasets (runs #2 to #7) are analyzed, in which signals are sampled at 20 kHz . For these runs we only have one photosensor signal, other than the voltage signal and the two currents signals. Arcs are identified as previously described, using a variable threshold, andwe use the input signal combination $[V_1, I_1 + I_2]$ that has proved to be the best performing one.

As a first step we look for a common configuration of the time and frequency windows m and p, that gives good results for all the datasets. Thus, for each dataset, we train a SVM for different configurations of m and p, and we find the one that gives the highest 5-fold CV. The grid-search approach is used to determine the SVM parameters, i.e. C and γ, with

an RBF kernel. Then we look for a fixed configuration of m and p that gives a 5-fold CV in all datasets that has the minimum mean squared error, MSE, with respect to the best CV values found previously. The tested values for the windows are $m = [400,800,1200]$ and $p = [30,36,42,48]$, different from the respective values in section 3.2 because of the different sampling time.

Table III

Dataset#	*m* / *p* relative to max CV	Max CV(%)	CV (%) for *m*=1200 and *p*=48
2	800 /30	91.0682	90.6705
3	800 /42	92.1136	90.5795
4	400/36	92.2045	95.3295
5	1200/48	94.1932	94.1932
6	400/30	89.7273	87.4318
7	800/42	90.2159	89.4432

The second column of Table II shows the values of *m* and *p* that give the max 5-fold CV (shown in the third column) for each test case, while the last column shows the CV for *m*=1200 and *p*=48, which is the case with minimum MSE with respect to the second column. From this result we can give the suggestion that using the largest windows, both in time and frequency, gives the best average results among several different runs.

This last result can be given as a guideline for the implementation of an SVM dedicated to the classification of such kind of signals. In addition, the accuracy obtained for the first run, are confirmed also for these additional runs, showing that the SVM based approach is suitable for this application.

3.5 Generalization of the SVM performances

All the results described in the previous subsections are relative to SVMs trained and tested on data coming from the same run. For the method to be useful in practice, we have to verify its robustness in terms of applicability when trained with data coming from a run and tested with data coming from a different run (possibly relative to different trains running on different tracks). In addition, there are 6 test runs sampled at a higher frequency than the first one; for this reason we have decided to consider the first run as the reference one.

According to what described above, the first experiment is related to the classification using down-sampled signals form 20 kHz to 5 kHz. To this purpose all the signals of runs #2 to #7 have been down-sampled to 5 kHz and for each case an SVM has been trained using the pre-processing parameters determined in the first case study: $m = 150$, $p = 12$.

Table IV

Dataset#	CV (%)
2	88.7045
3	89.2841
4	95.6928
5	90.1250
6	85.1705
7	88.3068

The 5-fold CV are shown in table IV: the selected time windows from the first case study give acceptable results also on the other six cases, using down- sampling. This is a good result under two point of views:

- downsampling the signals does not significantly affect accuracy
- The pre-processing parameters determined for a test run (train and track) perform very well also with different data without the need of repeating the parameters setup.

As an additional comment on the required sampling frequency, we can say that the arc influence on voltage and current, in the frequency domain, is significant in the lower frequencies up to hundred of Hz, so the use of higher sampling frequencies is not necessary.

The last numerical experiment investigates the use of an SVM model, trained with one dataset, to classify the arcs of other datasets, without retraining. The model trained in the first case study which gives the best CV value is here used to classify the datasets #2 to #7, that are used as test data in their entirety, after down-sampling to 5 kHz.

Table V

Dataset#	Accuracy (%) on the whole set down-sampled at 5kHz
1	84.6295
2	74.7682
3	78.2414
4	83.4477
5	76.1932

6	84.5523

Table V shows the results of the accuracy on the whole sets. These results show that a well trained SVM is able to generalize the arc classification capability also to signals from different trains in different trips and conditions.

4. Classification with radial basis function networks

Other supervised classification techniques may be used instead of SVM, for instance Radial Basis Function networks (RBF), also a classifier ensemble may be considered, by using some classifier fusion technique. The choice of SVM has been made by the authors as they represent one of the best performing classifiers, and with the aim of determining the classification accuracy achievable with a well known method for this industrial application. For the sake of completeness in this section the authors report the results obtained using an RBF classification approach [21].

RBF networks have typically three layers: an input layer, a hidden layer with a non-linear activation function and a linear output layer. Given an input $\mathbf{x}_i \in \Re^p$, $i = 1..2N$, the scalar output of the network is a given by the classification function:

$$\hat{y}_i = \sum_{k=1}^{N_C} a_k \exp\left(-\beta \left\| \mathrm{x}_i - \mathrm{c}_k \right\|^2\right) \tag{10}$$

where N_C is the number of neurons in the hidden layer, $\mathrm{c}_k \in \Re^c$ is the center vector for neuron k, β is the width of the Gaussian radial basis function, a_k is the weight of the neuron k in the linear output layer and $\|\|$ represents the Euclidean distance. The training is typically performed in two phases, first fixing the width and centers and then determining the weights. In the first unsupervised step, the center vectors c_k are chosen, and this step can be performed in several ways: centers can be randomly sampled as a subset of given points $\mathbf{x}_i$, or obtained by Orthogonal Least Square Learning Algorithm, or found by clustering the input samples and using the cluster centers. In this work we use the k-means clustering algorithm

for determining the centers c_k. The width β is usually fixed for all neurons to a value which is inversely proportional to the maximum distance $d_{\max}$ between the chosen centers. In this work we use $\beta = 1/(c_w d_{\max})$, where c_w is an integer constant to be determined. Given the center vectors and the width, the weights of the output layer $\mathrm{a} = [a_k]^T, k = 1 \ldots N_C$ are computed by a pseudoinverse solution:

$$\mathrm{a} = \mathrm{M}^{+}\mathrm{y} \tag{11}$$

where $\mathrm{M}(i,j) = \exp\left(-\beta \left\| \mathrm{x}_i - \mathrm{c}_j \right\|^2\right)$, $i = 1..N_{train}$, $j = 1 \ldots N_C$, and $\mathbf{y} \in \Re^{N_{train}}$ represents the known target vector with labels for training. In fact equation (10) may be written as $\hat{\mathrm{y}} = \mathrm{Ma}$ and (11) represents the unique minimizing solution of the training error function $\left\| \hat{\mathrm{y}} - \mathrm{y} \right\|$.

The hyper-parameters to be determined are the number of centers N_C and the constant c_w that determines the width. These are determined using an exhaustive search approach and 5-fold cross validation.

The test case used for comparison with SVM is the test run #1 analyzed in section 3.3. The same preprocessing parameters are used as in the case of SVM, (300 samples for time window and 12 samples for frequency window). In the RBF networks, as N_C grows the performance of the network always improves, so we use the following approach. The network is trained increasing N_C up to a value where 5-fold CV does not improve more than 0.01% (demonstrate a plateau). This is repeated for the following values of $c_w = 1,2,\ldots,10$. As a result we have found the following best RBF parameters: $N_C = 120$, $c_w = 6$, obtaining a 5-fold CV of 85%. The best 5-fold CV obtained with SVM, which is shown in table II, for the considered preprocessing parameters, is 92.66%, where SVM parameters have been determined using the two step grid search approach described above. From this comparison we may conclude that the SVM approach reveals to be superior to RBF for this particular application in the case considered.

A more detailed study of RBF and other classifiers, also in the optic of creating an

ensemble of classifiers, is considered a topic for future work and is out of the scope of this paper.

5. Conclusion

In this paper the authors propose a SVM based technique for classifying a set of experimental data recorded on a run of a high speed train.

The recorded data are voltage, current and the output of a phototube, showing when an arc event is detected. The classification algorithm tries to determine the presence of the arc by analyzing the voltages and currents, while a direct look to the data would not allow the location of the electric arc. The procedure gives good results, which means an accuracy of around 90% in the arc location; this value might seem too high if compared to the results in Table V, however, looking at a practical implementation of the procedure, the available data are relative to 7 test runs (electrical quantities and phototube), and it would make sense to use all of them to train a generalized SVM, leading to results similar to table IV.

After these first encouraging results, the research will be carried on to define the requirements of the signal acquisition and processing system and their compatibility with the equipment already installed on board. Subsequently more extensive test will be performed in order to define the condition based maintenance strategies.

Acknowledgment

This work was supported by the Italian Ministry of University (MIUR) under a Program for the Development of Research of National Interest (PRIN grant 20089J4SM9).

References

[1] Swift, M., Aurisicchio, G., Pace, P.: “New practices for railway condition monitoring and predictive analysis”, proceedings of the IET Conference on Railway Condition Monitoring and Non-Destructive Testing, RCM 2011, pp. 1 – 6.
[2] Betts, A. I., Hall, J. H., Keen, P. M.: "Condition monitoring of pantographs" Proceedings of International Conference on Main Line Railway Electrification, pp. 129 – 133, 1989.
[3] Collina, A., Fossati, F., Papi, M., Resta, F.: "Impact of overhead line irregularity on current collection and diagnostics based on the measurement of pantograph dynamics", Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, vol. 221, no. 4, pp. 547-559, 2007.
[4] Elia, M., Diana, G., Bocciolone, M., Bruni, S., Cheli, F., Collina, A., Resta, F.:

"Condition monitoring of the railway line and overhead equipment through onboard train measurements – an Italian experience", proceedings of the IET Conference on Railway Condition Monitoring, RCM 2006, pp. 102 – 107.
[5] Usuda, T., Ikeda, M., Yamashita, Y.: "Prediction of contact wire wear in high speed railways", proceedings of the 9th World Congress on Railway Research, 2011, pp. 1 – 10.
[6] Daadbin, A., Rosinski, J.: "Development, testing and implementation of the Pantograph Damage Assesment System (PANDAS), Computers in Railways XII, WIT Press, 2010, pp. 573 – 578.
[7] Jutard, M., Fitaire, M., Le Duc, E.: "Moyensd'étude des arcs de rupture du contact pantographe-cateénaire", Revue GénéraledeslChemin de Fer, vol. 108, no. 11, pp. 5 - 15, 1989.
[8] Bruno, O., Landi, A., Papi, M., Sani, L.:" Phototube sensor for monitoring the quality of current collection on overhead electrified railways", Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, vol. 215, no. 3, pp. 231-241, 2001.
[9] Landi, A., Menconi, L.,Sani, L.: "Hough transform and thermo-vision for monitoring pantograph-catenary system", Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, vol. 220, no. 4, pp. 435-447, 2006.
[10] Östlund, S., Gustafsson, A., Buhrkall, L., Skoglund, M.:"Condition Monitoring of Pantograph Contact Strip", proceedings of 3rd Railway Condition Monitoring Conference, Derby UK, June 2008.
[11] Huang, H. H., Chen T. H.: "Development of method for assessing the current collection performance of the overhead conductor rail systems used in electric railways" Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, vol. 222, no. 2, pp. 159-168, 2008.
[12] Barmada, S., Landi, A., Papi, M., Sani, L.: "Wavelet multi-resolution analysis for monitoring the occurrence of arcing on overhead electrified railways", Proc. Instn. Mech. Engrs. vol. 217 part. F: J. Rail and Rapid Transit, pp. 177 - 187, 2003.
[13] Barmada, S., Musolino, A., Tucci, M.: "Voltage and Current Analysis for Arcing Detection on AC Overhead Electrified Railways", Proceedings of Applied Computational Electromagnetics Society Symposium, ACES 2011, 27 - 31 March, 2011, Williamsburg, USA, pp. 975 - 979.
[14] Caiado, J., Crato, N., Peña, D.: "A periodogram-based metric for time series classification," Comput. Stat. and Data Analysis, vol 50, 2006, pp. 2668–2684
[15] Cortes, C., Vapnik, V.: "Support-vector network," Machine Learning, vol. 20, pp. 273-297, 1995.
[16] Boser, B. E., Guyon, I., Vapnik, V.: "A training algorithm for optimal margin classifiers," Proceedings of the Fifth Annual Workshop on Computational Learning Theory. ACM Press, pp. 144-152, 1992.
[17] Platt, J. C.: "Fast training of support vector machines using sequential minimal optimization,", Advances in Kernel Methods - Support Vector Learning, Cambridge, MA, 1998. MIT Press.
[18] Hsu, C. W., Chang, C. C., Lin, C. J.: " A practical guide to support vector classification", Tech. rep., Department of Computer Science, National Taiwan University, 2003.
[19] V. Cherkassky, Y. Ma "Practical selection of SVM parameters and noise estimation for SVM regression", Neural networks, 2004 - Elsevier, Volume 17, Issue 1, January 2004, Pages 113–126
[20] F. Friedrichs, C. Igel," Evolutionary tuning of multiple SVM parameters", Neurocomputing Volume 64, March 2005, Pages 107–117
[21] J Park, IW Sandberg " Universal approximation using radial-basis-function networks", Neural computation, 1991, Vol. 3, No. 2, Pages 246-257

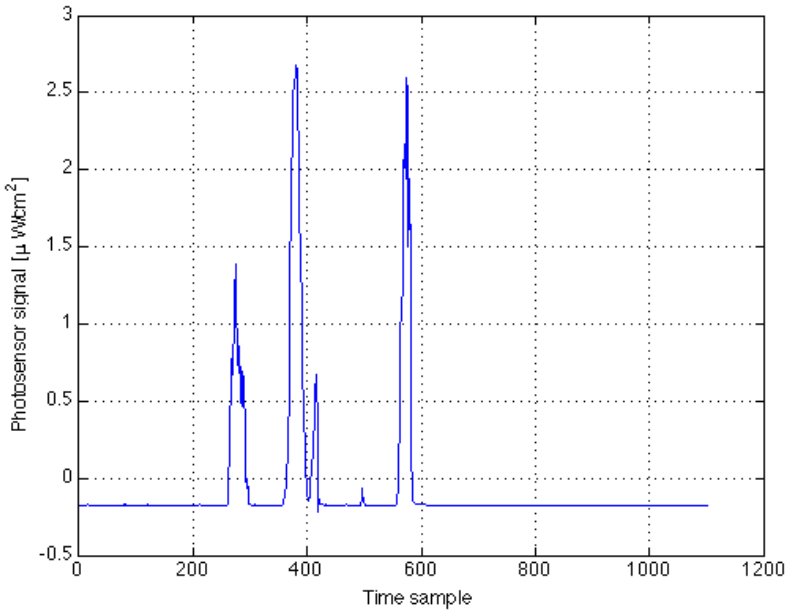

Fig. 1. Phototube output (run #1)

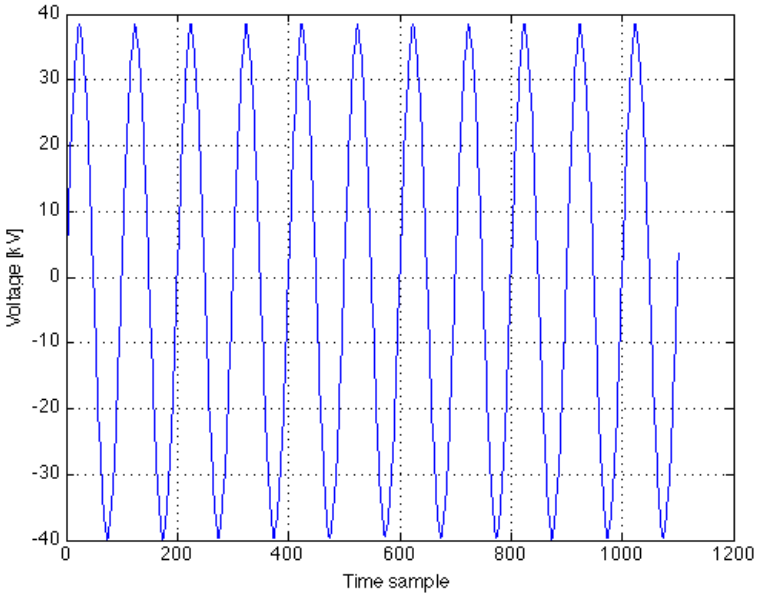

Fig. 2. Recorded voltage (run #1)

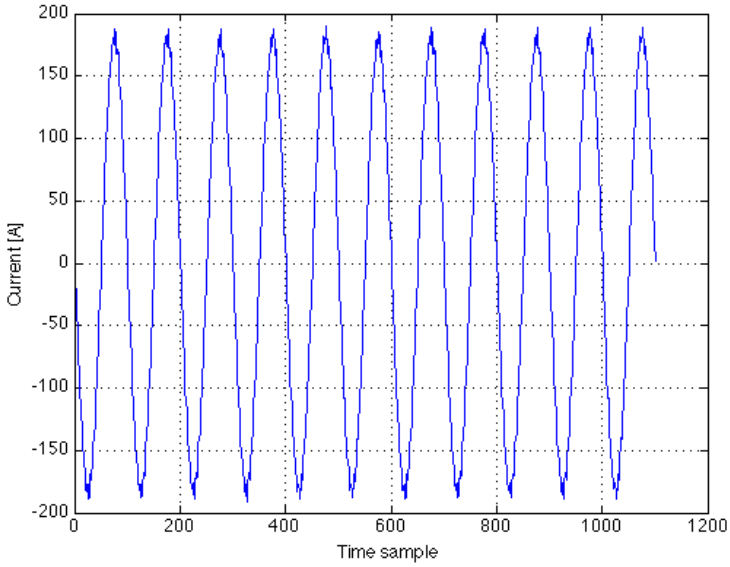

Fig. 3. Recorded current(run #1)

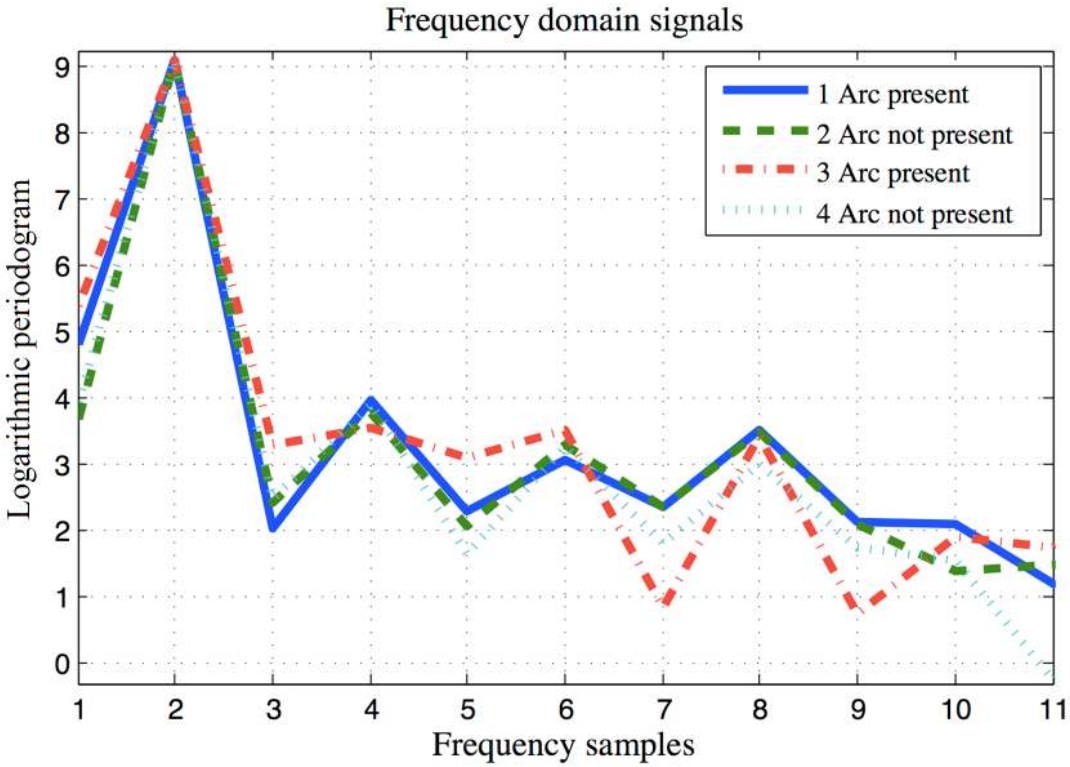

Fig. 4. Current periodogram (run #1)

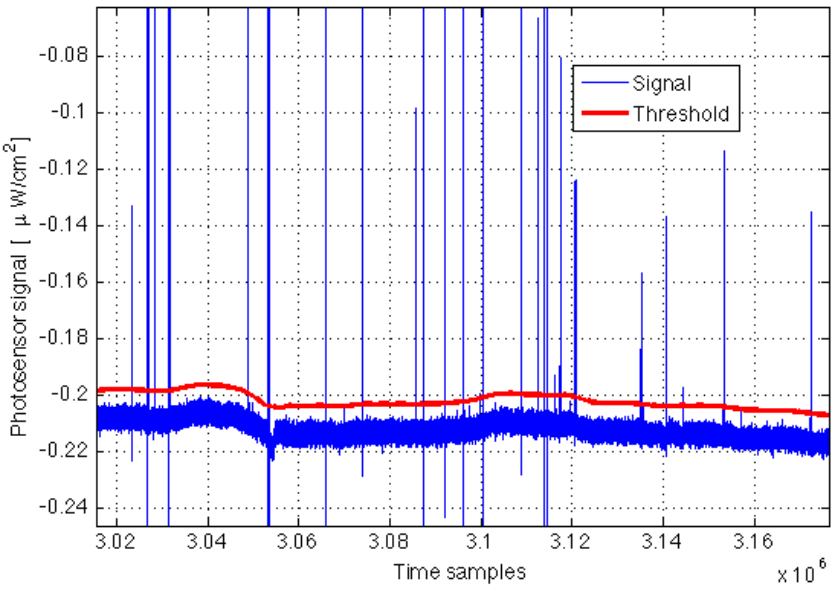

Fig. 5. Variable threshold on photosensor signal.

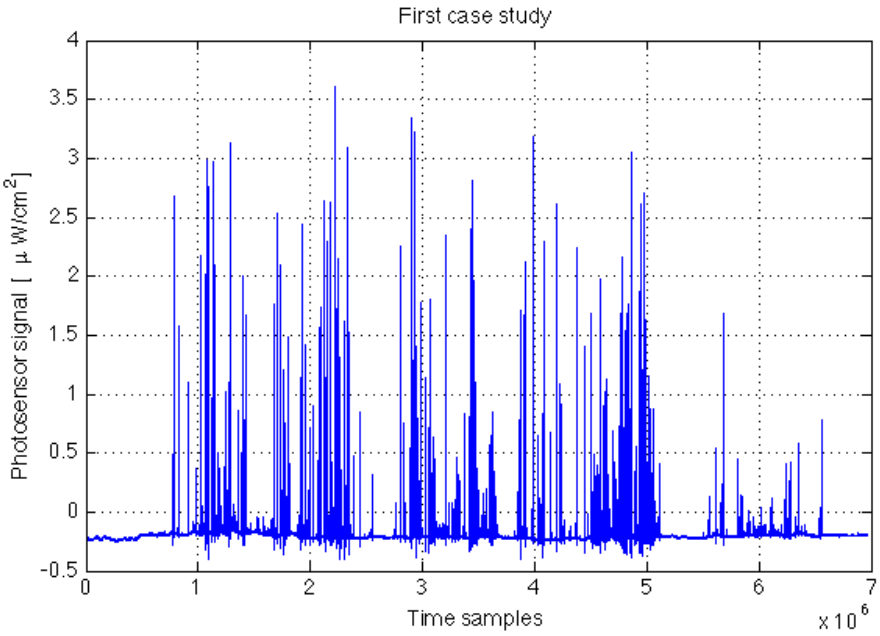

Fig. 6. Complete signal of one photosensor