EasyChair Preprint

№ 359

Selection of Variables by the F-Score Algorithm for Radiated Magnetic Field Signals Discrimination of Electrical Discharges

Mohamed Gueraichi, Hocine Moulai and Azzedine Nacer

July 20, 2018

Selection of Variables by the F-Score Algorithm for Radiated Magnetic Field Signals Discrimination of Electrical Discharges

Mohamed Gueraichi, Hocine Moulai and Azzedine Nacer
Laboratory of Industrial and Electrical Systems
University of Sciences and Technology Houari Boumediene (USTHB), 32, El Alia, Bab Ezzouar
16111, Algiers, Algeria
mgueraichi@usthb.dz, hmoulai@usthb.dz, anacer@usthb.dz

Abstract— **This paper proposes the use of a hybrid variables selection method called F-Score, which reduces the duration of the learning and test phases while improving the accuracy of recognition for radiated magnetic field signals discrimination of electrical discharges generally taking place in insulation systems and in particular in insulators of high-voltage lines and power transformers. The classifier used is based on support vector machines (SVMs). The experimental analysis was carried out on a basis of data comprising respectively 161 signals of which 106 learning and 55 for the test. The obtained results show that the proposed algorithm combined with the SVMs, allows a substantial reduction in the number of variables and a high improvement of the recognition rate compared to the pre-selection rate.**

Keywords- ***Electrical Discharges, Insulation systems, Support Vector Machines (SVMs), Variables Selection, F-Score.***

I. INTRODUCTION

We can distinguish two types of discharges: dangerous that can lead to a breakdown and consequently the partial or total destruction of the equipment, and not dangerous that extinguish themselves [1, 2]. The radiated magnetic field signals associated with these acquired electric discharges are represented by 25,000 variables. It is a question of discriminating the two types of discharges and mainly of focusing on the detection of the radiated magnetic field signals linked to dangerous type discharges. To do this, we use a recent classifier, namely support vector machines or SVMs, whose robustness in the field of pattern recognition has been proven [3, 4, 5, 6, 7], Moreover, unlike the neural networks [8, 9, 10], used in the past which are binding in terms of the number of parameters to be adjusted, SVMs require to adjust only two parameters, which is a great advantage.

The performance of the classifier strongly depends on the quality of representation of the examples or in our case of the signals to be recognized: harmless or dangerous, which generally implies the obligation to represent them by means of a large number of variables [11]. It is therefore common for some of these to contain only redundant, noisy or irrelevant information to the classification, thus making the learning of the recognition system more complex [11].
The selection of variables is a process that aims to filter the basic characteristic vector, so that only the discriminant information is extracted and presented to the classifier in a relevant way.

Several methods have been developed. We can divide them into two categories: Filtering methods (Filters) and enveloping methods (Wrapper) [3, 11, 12].

The Filter approach is performed as a preprocessing since the variables are selected independently of the classifier and regardless of its influence on system performance. This model is characterized by its speed. Nevertheless, the effectiveness of this type of method is questionable [3, 11, 12].
The Wrapper approach depends on the classifier used which contributes to the evaluation of the quality of the subset of selected variables. For each subset found, it is necessary to train a classifier until the empirical error is validated. This model is efficient but is characterized by a high calculation time [3, 11, 12].

We propose to use a combination of filter and wrapper methods using a Fisher-based measure called F-Score [15] that assigns a score for each variable, which allows the variables to be ranked by importance. Once the redundant variables are eliminated, we will search for the optimal percentage of the number of remaining variables sorted in decreasing order leading to the best recognition rate.

The rest of this article is organized as follows:

In Section II, we give a brief description of the SVMs classifiers, which are currently the best in terms of performance [3]. In Section III, we will analyze the selection of variables by the F-Score algorithm as it was designed and then describe in detail this algorithm applied to our system.

In section IV will be presented the experimental results and their interpretation.

Finally we conclude on the work done and the remarks that can be made.

II. SVMs Classifiers

Support vector machines or SVMs were developed in the 1990s by Vapnik [3] and supplanted other types of classifiers such as neural networks. They have become popular with the scientific community for almost two decades and have since become widely used, particularly in the field of pattern recognition. In fact we have a set $\{(x_i\ ,\ y_i)\}_{i=1}^{n}$ where $x_i \in R^m$ are the n examples of learning while m is the number of variables for each example, for their part, y_i are the labels corresponding to the two classes: $\{-1, +1\}$.

Originally, the purpose of the SVMs is to separate two classes of data, for this purpose we search for the best separation plan that maximizes the margin between the data of the two classes by calculating a decision function whose expression is the next one :

$$f(x) = sign(\sum_{k=1}^{S_v} \alpha_k\, y_k K(x_i, x_k) + b) \qquad (1)$$

Where α_k are the parameters of Lagrange, S_v is the number of support vectors x_k which are actually the border points of the margin such as $0 \leq \alpha_k \leq C$ where C is the regularization parameter to be adjusted that constitutes a trade-off between the maximization of the margin and the error due to the non-separable data. Finally x_i and b represent respectively the variable vector and the bias.

In practice, the data is nonlinearly separable, so a kernel function $K(x_i, x_k)$ is used.

The kernel function used is the RBF known for its better performance compared to other kernels [3, 4, 5, 6, 7] whose expression is:

$$RBF(x, x_k) = exp\left(-\frac{1}{2\,\sigma^2}\|(x - x_k\|^2\right)$$

Where σ is the kernel parameter. It will be a question of adjusting then of finding the optimal couple (σ, C) which allows to find the best recognition rate. In addition, the kernel function must satisfy the conditions of Mercer.

III. Selection of Variables Using Algorithm F-Score

Selection of variables is a method of choosing a subset of relevant variables that is optimal or sub-optimal from a set of original variables according to a criterion of evaluation.

The goal is to make a selection from the original feature vector $X_n = [x_1, x_2, \dots, x_n]'$ while trying to improve performance or failing to preserve them. The vector obtained after selection will therefore be:

$$X_m = [x_1, x_2, \dots, x_m]' \quad \text{with} \quad m < n$$

The authors set out a list of three objectives for making a selection of variables [3, 11]:

1/ Reducing the size of learning and test bases.

2/ The reduction of learning and recognition times.

3/ Improvement or failure to save the recognition rate.

The F-Score technique is used to calculate the discrimination between two classes by calculating a score called Fisher score from the learning data.

It is useful to specify that Fisher's criterion consists of calculating the distance between the average values of a variable established for two classes of data, and of normalizing it by the average of the variances, in order to estimate the discriminative power of the variable considered between these two classes [15].

This criterion is written for a given variable i :

$$F(i) = \frac{(\mu_i^{(1)} - \mu_i^{(2)})^2}{(\sigma_i^{(1)})^2 + (\sigma_i^{(2)})^2} \qquad (2)$$

For two classes designated by (1) et (2) whose number of examples are respectively n_1 and n_2 , this criterion $F(i)$ for the ith variable is defined by: [15]

$$F(i) = \frac{\left(\bar{x}_i^{(1)} - \bar{x}_i\right)^2 + \left(\bar{x}_i^{(2)} - \bar{x}_i\right)^2}{\frac{1}{n_1-1}\sum_{k=1}^{n_1}\left(x_{k,i}^{(1)} - \bar{x}_i^{(1)}\right)^2 + \frac{1}{n_2-1}\sum_{k=1}^{n_2}\left(x_{k,i}^{(2)} - \bar{x}_i^{(2)}\right)^2} \qquad (3)$$

Where $\bar{x}_i$ is the mean of the variables of rank i for the two classes (1) and (2) combined or what amounts to the same at the training base.

$\bar{x}_i^{(1)}$ and $\bar{x}_i^{(2)}$ are respectively the means of variables of rank i of classes (1) et (2).

$x_{k,i}^{(1)}$ and $x_{k,i}^{(2)}$ are the variables of the kth example of rank i respectively of each of the two classes (1) and (2).

The numerator indicates the discrimination between the two classes, while the denominator indicates the discrimination within each of the two classes. Thus, the higher the $F(i)$, the more the variable is discriminant [15].

This score is therefore used as a variable selection criterion in our problem of discriminating radiated magnetic field signals, from the expression (3) of the F-Score, we propose to select and retain only the most relevant variables, which will also allow elimination of the redundant variables i.e. those with the same values of the Fisher score, all this leads to a reduction in the size of the variables and consequently a reduced calculation time for both the learning and the test.

IV. Description of F-Score Algorithm

A. Progress of F-Score Algorithm

Knowing that the highest $F(i)$ score corresponds to the most discriminating variable, the F-Score algorithm consists of six steps described as follows:

- F-Score calculation for each variable i.
- Elimination of redundant variables.
- Sort scores of the remaining variables in decreasing order from the most discriminating variable to the least discriminating variable.
- Choice of several thresholds for the F-Score.
- Calculation of the rate of recognition of dangerous signals for each chosen threshold.
- Retention of the threshold corresponding to the best recognition rate.

B. Choice of Threshold

The difficulty of the F-Score method lies in the choice of a threshold that can determine the optimal subset of the variables to be selected.

Several empirical thresholds must be used to find the best approach for choosing the F-Score threshold to give the best recognition rate.

C. Prioritization of Fisher's Scores

The approach for which we have opted was to take as a threshold a variable percentage of the number of variables (amputated redundant variables) from the variable with the best F-score to the one with the lowest and to determine experimentally the optimal threshold leading to the subset of variables corresponding to the best recognition rate.

V. Experimental Results

We present in this section an experimental analysis of the performances of the F-Score method in the selection of variables for the discrimination of radiated field signals. To do this, we proceeded in two stages. The first stage is to perform the recognition of these signals, using the original signals i.e. without selection of variables; classification is done using support vector machines (SVMs).

The second stage is to carry out the recognition after implementation of the F-Score algorithm on the original signals which allows to obtain vectors of reduced dimension, where the number of variables has become much smaller.

We looked at 161 examples for the database we generated at our laboratory, including 106 examples for learning and 55 for testing.

The classes of the harmless type and dangerous type signals are formed respectively of 100 and 61 signals.

We note that our database is formed mainly of signals associated with discharges of low voltages and not dangerous while the signals associated with discharges of high voltages and dangerous are a minority. Signals numbered from 1 to 100 consist of non-dangerous signals and constitute the first class while those numbered from 101 to 161 are dangerous and constitute the second class.

We have taken $2/3$ of the database of each class of signals for learning and $1/3$ remaining were reserved for testing.

As an illustration, for the 106 examples of the learning base, the values of the F-Score as a function of the rank i of the variable are as follows:

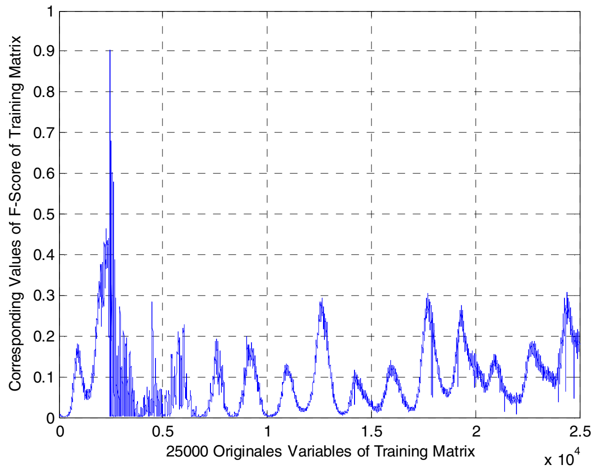

Figure.1: Score F (i) according to the rank of the variable i

A careful examination of the curve in Figure 1 shows that there is a large variation in the values of some variables. Some have the same score: these are the redundant variables, which led us to keep only one of them having an identical F-Score value.

Finally, the selection process that we have adopted takes place in four stages:

- Sort the original F-Score vector that contains the 25000 starting variables in descending order.
- Keep only one of the variables with the same $F(i)$ score, since they provide the same information.
- Sweep the new reduced F-Score vector and sort it in descending order by taking percentages of the number of variables ranging from 5% to 95% in 5% steps and observe the recognition rate for each percentage.
- Retain the percentage of variables that gives the best recognition rate, or if not the same before selection.

Figure.2 shows the evolution of the dangerous signal recognition rate of test designated by DSTEST, as a function of the percentage of the number of variables, from the most discriminating to the least discriminating.

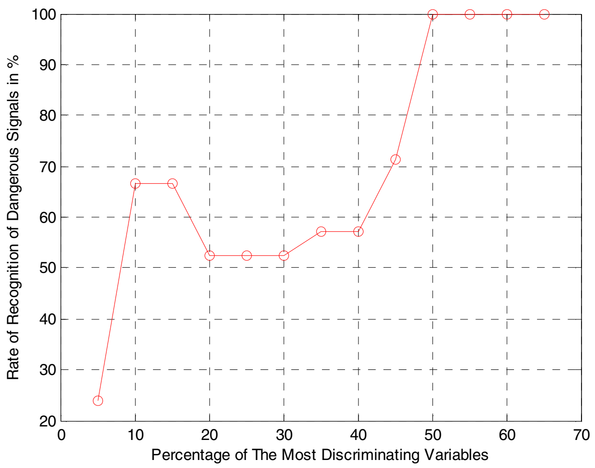

Figure.2: Rate of recognition of dangerous signal after selection according to the percentage of the most discriminating variables

From Figure 2, we find that the best recognition rate after selecting variables is 100%, which corresponds to several percentages corresponding to 50, 55, 60 and 65% respectively.

The threshold of 50% of total number of no-redundant variables corresponds to the optimal threshold because it corresponds to the minimum of variables, i.e. exactly 1178 discriminant variables selected, the pair (C, σ) is determined experimentally as (471,11).

In total, out of 25,000 variables, a total of 22,645 redundant variables were eliminated while on the remaining variables, only 1178 variables were retained, that is to say almost 4.7% of all the original variables.

In TABLE.I below, the DSTEST recognition rate, the duration of learning and test phases, and the number of variables used before and after the application of F-Score selection are summarized respectively.

TABLE I. RECOGNITION RATE DSTEST, DURATION OF LEARNING AND TEST PHASES AND NUMBER OF VARIABLES

Recognition Rate DSTEST before and after application of the F-Score Algorithm	Duration of Learning and Test phases	Number of Variables
With only SVMs : 28.57%	20h 26' 40''	25000
F-Score + SVMs : 100 %	3h 35' 8''	1178

VI. CONCLUSION

This article is an alternative hybrid approach to the filter and wrapper models using two selection criteria, one independent of the classifier (the $F(i)$ score), while the second criterion corresponds to the recognition rate for dangerous signals in insulation systems.

Before the application of the F-Score selection algorithm, the SVMs alone were able to recognize only 6 of the 21 dangerous type test signals, which is a low recognition rate of 28.57%. The presence of a very large number of redundant variables numbering 22645 that provide no additional information to the classifier and also the non-redundant but irrelevant variables that may be for some noisy.

To solve this problem, we made a selection of variables; we opted for the application of the F-Score algorithm before classifying the two classes of signals by the SVMs.

The strategy of the threshold proposed above, had very satisfactory results: after elimination of 22645 redundant variables, only 2355 variables underwent a selection process, among them, only 1178 were selected and retained corresponding to an optimal threshold equal to 50% of the proposed percentage.

The F-Score algorithm combined with the SVMs not only reduced the number of variables very significantly but also led to the recognition of all dangerous type signals (SDTEST = 100%) with a significantly reduced computation time for the phases of learning and testing, almost six times less than pre-selection.

REFERENCES

[1] H. Moulai, "Etude des Courants de Préclaquage dans les Diélectriques Liquides," Doctorat d'Etat Thesis, The National Polytechnic School, Algiers, Algeria, 2001.

[2] F. Aberkane, "Etude des Processus de Décharges Electriques dans les Diélectriques Liquides," Doctorat Thesis, University of Sciences and Technology Houari Boumèdiene, Algiers, Algéria, 2015.

[3] M. Cheriet, N. Kharma, C.L. Liu and C. Y. Suen, Characters recognition systems. A Guide for students and Practionner. Edited by J. Wiley , 2007, pp. 90-103.

[4] S. Theodoridis and K. Koutroumbas, Pattern Recognition, Fourth Edition, Edited by Academic Press , Elsevier Inc, 2009, pp 261-318.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp. 68–73.

[5] R.O.Duda, P.E.Hart and D.G.Stork, Pattern Classification, snd Ed, pp 114-115.

[6] J.P. Marques DE S, Pattern Recognition, Concepts, Methods and Applications. Edited by Springer, 2001, pp. 88-89.

[7] N.E. Ayat, ''Sélection de modèles automatique de machines à vecteur de support : Application à la reconnaissance d'images de chiffres manuscrits,'' Doctorat Thesis, Ecole de Technologie Supérieure, Quebec.

[8] F. Aberkane, A. Nacer, H. Moulai, F. Benyahia, and A. Beroual, "ANN and Multilnear Regesssion Line Based Discrimination Technique between Discharge currents for Power Transformers Diagnosis," 2ndInternational Advances in Applied Physics and Material Science Congress, April 26-29 2012, Antalya, Turkey, pp. 1-5.

[9] F. Aberkane, H. Moulai, A. Nacer, F. Benyahia, and A. Beroual, "ANN and Wavelet Based Discrimination Technique between Discharge currents in Transformers Mineral Oil," The European Physical Journal Applied Physics, Vol 58, N° 2, 2012, pp. 1-6.

[10] A. Schenk, "Surveillance Continue des Transformateurs de Puissance par Réseaux de Neurones Auto-Organiséés," Doctorat Thesis, Ecole Polytechnique Fédérale de Lausanne, 2001.

[11] F. Grandidier,''Un Nouvel Algorithm de Sélection de Caractéristiques: Application à la lecture automatique de l'écriture manuscrite,'', Doctorat Thesis, Ecole Supérieure de Technologie, Quebec, 2007, pp. 136-140.

[12] H. Liu, H. Motoda,''Computational Methods of Feature Selection,'' Edition Chapman & Hall/CRC. pp. 25-28.

[13] S H. Chouaib. Tabbone, O. R. Terrades, F.Cloppet et N. Vincent, "Sélection de caractéristiques à partir d'un algorithme génétique et d'une combinaison de classifieurs Adaboost,", Actes du dixième colloque international Francophone sur l'écrit et le document, October 8, 2008, Rouen, France, pp. 181-186.

[14] S. Guérif et Y. Bennani. 2007. "Sélection de variables en Apprentissage Numérique Non Supervisée,",Conférence Francophone sur L'apprentissage Automatique (CAP'2007), Juillet 2007, Grenoble, France, pp. 221-236.

[15] Y. W. Chen and C. J Lin, ”Combining SVMs with various feature selection strategies,” In feature Extraction, Foundations and Applications, Edited by Springer-Verlag, Berlin; 2006.

[16] X. Zheng, Y.W. Chen, C. Tao, D. Van Alphen, ”Feature Selection using Recursive Feature elimination for Handwritten Digit Recognition,”, IIH.MSP’09. Fifth International Conference on Information Hiding and Multimedia Signal Processing, September 2009, Washington, DC, USA, pp. 1205-1208.

[17] P. Zhang, T.D. Bui, C.Y. Suen, ” Hybrid Feature Extraction and Feature Selection for Improving Recognition accuracy Handwritten Numerals,”, Document Analysis and Recognition, 2007, pp. 136-140.