EFDA–JET–CP(10)03/15

S. González, J. Vega, A. Murari, A. Pereira, J.M. Ramírez,
S. Dormido-Canto and JET EFDA contributors

SVM-Based Feature Extractor for L/H Transitions in JET

SVM-Based Feature Extractor for L/H Transitions in JET

S. González[1], J. Vega[1], A. Murari[2], A. Pereira[1], J.M. Ramírez[3], S. Dormido-Canto[3] and JET EFDA contributors*

JET-EFDA, Culham Science Centre, OX14 3DB, Abingdon, UK

[1]Asociación EURATOM/CIEMAT para Fusión, Madrid, Spain
[2]Consorzio RFX – Associazione EURATOM ENEA per la Fusione, Padova, Italy
[3]Departamento de Informática y Automática UNED, Madrid, Spain
** See annex of F. Romanelli et al, "Overview of JET Results",*
(Proc. 22nd IAEA Fusion Energy Conference, Geneva, Switzerland (2008)).

Preprint of Paper to be submitted for publication in Proceedings of the 18th High Temperature Plasma Diagnostics, Wildwood, New Jersey, USA.
(16th May 2010 - 20th May 2010)

ABSTRACT.
Support Vector Machines (SVM) is a machine learning tool, originally developed in the field of artificial intelligence, to perform both classification and regression. In this paper we show how SVM can be used to determine the most relevant quantities to characterize the confinement transition from low to high confinement regimes in Tokamak plasmas. A set of 27 signals is used as starting point. The signals are discarded one by one until an optimal number of relevant waveforms is reached, which is the best tradeoff between keeping a limited number of quantities and not loosing essential information. The method has been applied to a database of 749 JET discharges and an additional database of 150 JET discharges has been used to test the results obtained.

1. INTRODUCTION

The aim of this work is to develop an automatic feature extractor to find the most relevant signals to characterize a physical behavior in large datasets. The system has been applied to study the transition from the Low (L) to the High (H) confinement mode (L/H transition).

The Feature Extractor System (FES) has been based on the development of a binary classification system to distinguish between L mode and H mode. This means that a pattern recognition system to determine the confinement regime has been developed. This is accomplished through the creation of a decision function to split the input space in two regions. This decision function has been used to find the main quantities that best describe the L/H transition.

27 signals (Table I) have been selected as candidates to provide discriminant characteristics for the pattern recognition problem. They include both physical quantities describing the plasma and geometrical parameters to take into account the position/shape of the plasma inside the vacuum vessel. This set of signals has been used in previous works about L/H transitions in JET [1].

A symmetric temporal segment of 100ms around the transition instant has been selected. This temporal length has been chosen because the best characterization of the transitions requires to be close enough to the change of regime. A typical temporal interval can be the one selected [1].

To develop the FES a set of 749 JET discharges has been used. The high dimensionality of the problem to solve should be noticed. The 27 signals are sampled with a period of 1 ms in the temporal segment around the transition. Therefore, each time instant is represented by a vector of 27 components (one per signal). This implies the use of 200 vectors per discharge which means to manage 149800 of these vectors to determine the decision function.

A first alternative to determine the quantities more relevant to characterize the L-H transition are tree structured methodologies, such as Classification and Regression Trees [2] (CART). The CART provides as result the variable ranking of the most relevant signals for a classification problem. CART is a non linear technique but presents the disadvantage of being very heavy in terms of computational times. Already for less than 100 discharges, the computation becomes prohibitively long.

A second alternative has been the use of the Principal Components Analysis (PCA) [3]. PCA is a method for reexpressing multivariate data. PCA creates a set of orthogonal components that are linear combinations of the original data with maximum variance. By examining the relationship between the Principal Components and the input data (Principal Components Loadings), it is possible to determine the importance of the original variables and eliminate the less interesting ones. This is carried out through their correlation matrix.

Unfortunately, the signals used to characterize the L/H transition show a small correlation, and thus, the variance covered by the first principal components is small (figure 1). For example, using the first eight principal components, only 84.51% of the original variance is covered.

The third option has been to use SVM [4] as classifier. This approach has the extra advantage of being able to easily handle the big amount of data required. The developed system is a feature extractor system (FES) that selects the most relevant characteristic for the L/H transition phenomenon.

This paper is structured as follows: an introduction to SVM is given in Section III. Section IV describes the feature extractor. Finally, the results of the JET L/H transition feature extraction are given in Section 4.

2. SVM FOR BINARY CLASSIFICATION

This section contains a brief explanation about SVM theory for binary classification problems.

For classification of nonseparable data, a function (called decision function) is obtained. This function divides the input space in two regions that classify the data into two different classes. The decision function is given by [6]:

$$D(x) = \sum_{i=1}^{n} \alpha_i^* y_i H(x_i, x_j) + b \qquad (1)$$

The yi are the labels of the respective classes. In the case of a binary problem, both classes can be distinguished by labels +1 and -1 respectively. The parameters α_i^*, $i=1,\ldots,n$ are the solution for the following quadratic optimization problem:

Maximize the functional:

$$L(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\; y_i \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j H(x_i, x_j)$$

subject to the constraints:

$$\sum_{i=1}^{n} y_i \alpha_i = 0; \qquad 0 \le \alpha_i \le C/n;\; i = 1,\ldots,n$$

given the training data (x_i, y_i), $i=1,\ldots,n$, an inner product kernel H, and a regularization parameter C.

Only a subset of the parameters α_i in equation (1) is nonzero. The data points x_k associated with the nonzero α_i are called Support Vectors (SVs). Therefore, the decision function is actually:

$$D(x) = \sum_{i=1}^{n} \alpha_i y_i H(x_i, x) + b \qquad (2)$$

The complexity of the model obtained is dependent on the number of SVs: the fewer SVs the lower the model complexity.

For a two class classification problem, SVM returns the separation hyperplane that maximizes the distance to the samples of both classes (figure 2).

3. SVM-BASED FEATURE EXTRACTOR

The objective of the SVM-based feature extractor is to find the most important features (signals) to characterize the L/H transition. It is a 2-class classification problem based on SVM.

The method consists of determining the relative weight among signals in the decision function. To this end, a separating hyperplane between the L and H confinement modes is computed with a linear kernel (figure 2 is an example of a separating hyperplane between two classes).

The hyperplane is a linear function that can be written as:

$$C_1X_1 = C_2X_2 + \ldots + C_mX_m + b = 0 \tag{3}$$

C_i: Weight of each feature
m: Number of features
b: Bias

It is possible to obtain the equation of this hyperplane by using the decision function given by equation (2). Taking into account that a linear kernel has the mathematical expression $H(x,x) = [x'E_i + 1]$, this decision function is:

$$D(x) = \sum_{i=1}^{n} \alpha_i\, y_i\, [x \cdot x_i + 1]$$

$$D(x) = \sum_{i=1}^{SV} \alpha_i\, y_i\, [(x_{i,1} \cdot x_1, \ldots, x_{i,m} \cdot x_m) + 1]$$

$$D(x) = \sum_{i=1}^{SV} (\alpha_i\, y_i x_{i,1} \cdot x_1) + \ldots$$
$$+ \sum_{i=1}^{SV} (\alpha_i\, y_i x_{i,m} \cdot x_m) + \sum_{i=1}^{SV} \alpha_i\, y_i$$

The decision function is equal to zero on the hyperplane.

Therefore,

$$0 = \left(\sum_{i=1}^{SV} \alpha_i\, y_i x_{i,1}\right) \cdot x_1 + \ldots$$
$$+ \left(\sum_{i=1}^{SV} \alpha_i\, y_i x_{i,m}\right) \cdot x_m + \sum_{i=1}^{SV} \alpha_i\, y_i \text{ , where} \tag{4}$$
$$C_j = \sum_{i=1}^{SV} \alpha_i y_i x_{i,j}, \quad b = \sum_{i=1}^{SV} \alpha_i y_i$$

Once the hyperplane equation has been obtained from the data, the FES must discard the least relevant feature. In the general equation of a hyperplane (3), the most important features are those whose C_k are the largest. Thus, the FES discards the smallest weight C_k and the process is repeated again with a feature space of dimension *m-1*. For example, the equation of the 2-dimensional hyperplane in figure 2 is $8x + 3y - 40 = 0$. Therefore, the most important feature of the training data is x. It can be proven that it is possible to choose a hyperplane $C_k x + b = 0$ that divides the example's feature space in two classes with less errors than a hyperplane $C_j y + b = 0$.

Figure 3 shows that the number of features in the final set relies on the model complexity of the hyperplane computed for each set of features (figure 3). The complexity of a hyperplane is measured by the number of SVs needed to define that hyperplane (equation 2). The number of features in the final set is the one where the number of SVs increases significantly from a set of m features to a set of $m-1$ features. In the case of figure 3, the final set of features contains 8 signals because it is the last set where the number of the SVs to define the separating hyperplane keeps the same tendency.

4. L/H TRANSITION FEATURE EXTRACTION

The main features for the JET L/H transition have been obtained from a starting set of 27 signals (Table I) in the interval of ±100ms around the transition and 749 JET discharges. The original signals have been resampled (1 ksamples/s) and normalized (to avoid scaling problems). Three different typical normalizations of the original data have been tested. In the first one, all data are rescaled to the interval [0, 1]. In the second one, the data are transformed to have mean 0 and variance 1. In the third normalization, the mean value of the signals is subtracted in each sample and then the data are rescaled to the interval [-1, 1].

The best classification success rate (92.20%) have been obtained using the normalization between 0 and 1.

The equation of the hyperplane, with the most important features from Table 1, is shown in Table 2. The success rate with this hyperplane has been tested with an additional set of 150 JET discharges. This test set provides a non-biased estimation of the accuracy of the results to avoid the potential overestimation that can affect the results of the training set. The success rate with these 150 discharges has been 90.04 %.

The set of the most relevant physical quantities to characterize the L/H transition in JET can be also validated with the use of non-linear kernels. For example, using a Radial-Basis Function (RBF) kernel, the classification success rate of the test set is 92.85 %: therefore the nonlinearities are not a dominating effect and the analysis performed with the linear kernel can be considered accurate.

It should be noted that the use of 200 samples/discharge, 749 discharges and feature vectors with high number of components (between 27 and 2) defines a very time consuming computation problem. Therefore, high performance computing has been necessary. A parallel version of SVM [7] has been used to allow the execution of the SVM classification problems. The computations have been performed in the Lince cluster at CIEMAT. It is made up of nodes Xeon 3.2Ghz with 2GB of RAM and a low-latency Infiniband network.

REFERENCES

[1]. J. Vega, A. Murari, G. Vagliasindi, G.A. Ratt. and JET- EFDA contributors, Automated estimation of L/H transition times at JET by combining Bayesian statistics and Support Vector Machines, Nuclear Fusion **49** (2009) 085023 (11pp)

[23]. Breiman L., Friedman J.H., Olshen R.A. and Stone C.J. 1993 Classification and Regression Trees (Belmont, CA/New York: Wadsworth Inc./Chapman and Hall)

[3]. J. Lattin, J.D. Carroll, P.E. Green, Analyzing Multivariate Data, 2003, Thomson Brooks/Cole

[4]. V.N. Vapnik, Statistical Learning Theory, John Wiley and Sons, Inc., 1998

[5]. V. Cherkassky, F. Mulier, Learning from Data : Concepts, Theory and Methods, 2007, John Wiley & Sons, Second Edition

[6]. J. Ramírez, S. Dormido-Canto, J. Vega and JET-EFDA Contributors, Automatic parallelization of classification systems based on Support Vector Machines: comparison and application to JET database, Fusion Engineering and Design (in press).

F1. Beta normalized with respect to the diamagnetic energy	**F15.** Time derivative of diamagnetic energy
F2. R coordinate inner lower strike point	**F16.** Z coordinate inner lower strike point
F3. R coordinate outer lower strike point	**F17.** Z coordinate outer lower strike point
F4. Electron temperature at PSI = 0.95 surface	**F18.** Electron temperature gradient at PSI = 0.95 surface
F5. Toroidal magnetic field at PSI = 0.95 surface	**F19.** Magnetohydrodynamic energy
F6. Outer interferometry channel	**F20.** R coordinate lower X point
F7. Upper triangularity	**F21.** Total heating power
F8. Z coordinate lower X point	**F22.** Safety factor at PSI = 0.95 surface
F9. Toroidal magnetic field	**F23.** Plasma current
F10. Radial inner gap	**F24.** Radial outer gap
F11. D_α view	**F25.** Lower triangularity
F12. Top outer gap	**F26.** Radiated power
F13. Line integrated density	**F27.** Elongation boundary
F14. Plasma inductance	

Table 1: Initial set of signals for the L/H transition.

FEATURE	ORIGINAL DATA	NORMALIZED DATA
F5. Toroidal magnetic field at PSI = 0.95 surface	-6.925446×10^{1}	-93.652816
F9. Toroidal magnetic field	5.457961×10^{1}	95.530676
F11. D_α view	-2.7359×10^{-16}	-23.191389
F13. Line integrated density	4.423290×10^{-2}	10.037411
F14. Plasma inductance	-8.888935	-12.933389
F19. Magnetohydrodynamic energy	6.333208×10^{-6}	27.013612
F23. Plasma current	2.850196×10^{-6}	8.884307
F26. Radiated power	-2.36697×10^{-6}	-109.091167
Bias	1.163064×10^{1}	-3.073230

Table 2: Coefficients of the hyperplane of the final set of signals

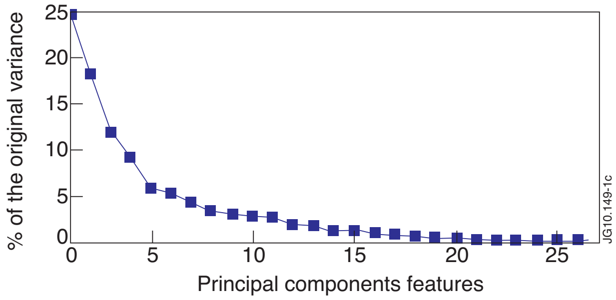

Figure 1: Principal Components Analysis of the L/H transition.

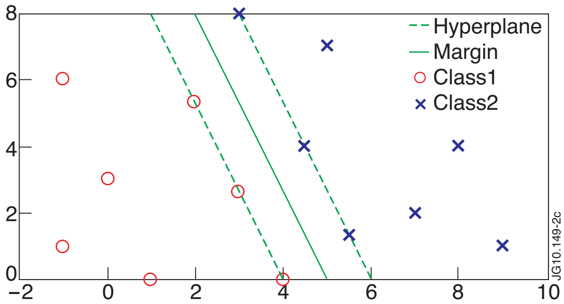

Figure 2: SVM classification in a two-dimensional space.

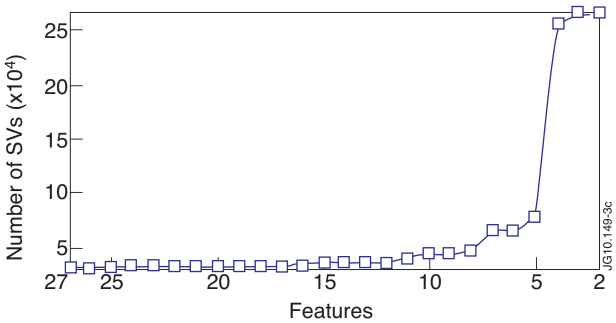

Figure 3: Example of the criterion used to select the final numberof features in the case study.