Classifier based on support vector machine for JET plasma configurations[a)]

S. Dormido-Canto,[1] G. Farias,[1] J. Vega,[2] R. Dormido,[1] J. Sánchez,[1] N. Duro,[1] H. Vargas,[1] A. Murari,[3] and JET-EFDA Contributors[4,b),c)]
[1]*Departamento de Informática y Automática, UNED, C/Juan del Rosal 16, 5a. 28040 Madrid, Spain*
[2]*Asociación EURATOM/CIEMAT para FUSIÓN, Avda. Complutense 22. 28040 Madrid, Spain*
[3]*Consorzio RFX-Associazione EURATOM ENEA per la Fusione. I-35127 Padua, Italy*
[4]*JET-EFDA, Culham Science Centre, OX14 3DB Abingdon, United Kingdom*

(Presented 14 May 2008; received 7 May 2008; accepted 14 July 2008; published online 31 October 2008)

The last flux surface can be used to identify the plasma configuration of discharges. For automated recognition of JET configurations, a learning system based on support vector machines has been developed. Each configuration is described by 12 geometrical parameters. A multiclass system has been developed by means of the one-versus-the-rest approach. Results with eight simultaneous classes (plasma configurations) show a success rate close to 100%. © *2008 American Institute of Physics.* [DOI: 10.1063/1.2972023]

I. INTRODUCTION

The shape of the last flux surface is an essential ingredient in the definition of the JET operation scenarios and several ones can be present during a discharge. Some kinds of data analysis are sensitive to the plasma configuration (for example, to the location of the *X*-point and strike points) and, therefore, proper identification (classification) of the plasma configuration is important.

At present, JET configurations are primarily identified by referring to an identifying keyword describing the request made, prior to the pulse, to the plasma control system. This has the disadvantage of being nonspecific, as several different identifiers can refer to the same configuration; cumbersome, as this data cannot be accessed automatically; incomplete, as some discharges are not assigned an identifier; and potentially wrong, as the identifier describes the intended rather than the resulting configuration. These problems motivated the development of an automated classifier.

Developing classifiers is a learning problem. It means that identification of different classes is needed to show the grouping of the data. The clustering can be carried out by exploiting *a priori* known information. This is known as supervised learning. Otherwise, if the data clustering does not use any prior information it is called an unsupervised classification method.

This article describes the development of a classification system for JET plasma configurations. It is a supervised system based on support vector machines (SVMs).[1] Section II reviews SVM as a technique for classification problems. It should be noted that any particular application of classification with SVM must address three key phases: feature extractions, training, and testing. These three phases for the present classification system are described in Sec. III. Section IV shows results for two different classifiers: one with three classes and one with eight classes systems. Section V is devoted to the final discussion.

II. SUPPORT VECTOR MACHINE FOR CLASSIFICATION

SVM is a universal constructive learning procedure based on statistical learning theory. SVM maps input data into a high-dimensional space using a nonlinear function. Once input data are mapped into the high-dimensional space, linear functions with constraints on complexity (i.e., hyperplanes) are used to discriminate the inputs, and a quadratic optimization problem must be solved to determine the parameters of these functions. Nevertheless for high-dimensional feature spaces, the large number of parameters makes this problem intractable. For this reason, duality theory of optimization is used in SVM to make the estimation of parameters in the high-dimensional feature space computationally affordable. The process time to classify in SVM is fast due to the matrix calculus in the algorithm.[2]

The linear approximation function corresponding to the solution of the dual problem is given in the kernel representation, $K(x,x')$, and it is called the optimal separating hyperplane. $K(x,x')$ represents a dot product of feature vectors in some high-dimensional space.[3] The solution in the kernel representation is written as a weighted sum of the support vectors, that is, a subset of the training data.

III. STAGES IN A CLASSIFICATION SYSTEM

A. Feature extraction

The description of signals in fusion databases is difficult to implement because there is no general solution for extracting generic features. So, feature extractors must be developed to extract the domain specific features most suited to

[a)]Contributed paper, published as part of the Proceedings of the 17th Topical Conference on High-Temperature Plasma Diagnostics, Albuquerque, New Mexico, May 2008.
[b)]For a full listing of names and affiliations of the JET-EFDA Contributors, see A. T. Macrander, Rev. Sci. Instrum. **79**, 10F701 (2008).
[c)]See the Appendix of M.L.Watkins *et al.*, Fusion Energy 2006 (Proceedings of the 21st International Conference, Chengdu, 2006) IAEA, (2006).

 © 2008 American Institute of Physics

Downloaded 11 Aug 2011 to 194.81.223.66. Redistribution subject to AIP license or copyright; see http://rsi.aip.org/about/rights_and_permissions

TABLE I. Geometrical parameters of the boundary of the last flux plasma surface.

Parameters	Description	Parameters	Description
elon	Elongation boundary	r_{og}	Radial outer gap
r_{geo}	Major radius	r_{ig}	Radial inner gap
r_{xpl}	Radial coordinate lower X-point	r_{mag}	Magnetic axis r coordinate
z_{xpl}	z coordinate lower X-point	z_{mag}	Magnetic axis z Coordinate
tri_l	Lower triangularity	el_{ax}	Elongation at magnetic axis
tri_u	Upper triangularity	vol_m	Plasma volume

the subsequent classification task. In the present case, the boundary of the last flux surface can be used to identify the plasma configuration of discharges. Therefore, geometrical parameters of the boundary have been chosen as feature vectors (Table I). These parameters were proposed by the experts.

B. Training and testing stages

The process of using data to determine the classifier is referred to as "training" of the classifier. Test data allow performing the "evaluation" of the classification system. Evaluation is important both to measure the performance of the system and to identify the need for improvements in its components (for instance, to add new attributes to the feature vectors or to eliminate redundant ones).

IV. CLASSIFICATION SYSTEMS

Results with two classification systems based on geometrical parameters of the last flux surface and SVM are presented in this section. In order to evaluate the approach, two different classifiers have been applied to some of the configurations stored in the JET database. These configurations belong to one of the following classes: HIXR_GB, StandardFat, HC_SFE_LT, SEPTUM, VLPC_SWEEP, D1F_C_SFE_LT, V_LFE_LT, D1Z_XFORM, and VH_3M5_HT3.

The first classifier has been trained to discriminate discharges belonging to the following three classes: VH_3M5_HT, HIXR_GB, and SEPTUM. Each configuration is defined by 12 geometrical parameters (Table I). Figure 1 shows that a simple visual inspection of parameters can be enough to solve the easiest cases. For the three above configurations, it is possible to discard parameters that do not contribute to the discrimination and to identify the ones that provide the most discriminative power: r_{ig} and z_{mag}. Therefore, for the three classes problem (VH_3M5_HT, HIXR_GB, and SEPTUM), two-dimensional feature vectors (parameters r_{ig} and z_{mag}) perfectly identify three clusters that correspond to the three different configurations (Fig. 2).

Figure 2 also shows three straight lines separating each class from all the others (linear discriminant functions). In this case, using a SVM classifier, in its simplest linear version, we obtain the hyperplane that maximizes the margin of separation, therefore minimizing the misclassification risk.

The data set for the three class problem is made of 199, 39, and 17 configurations, respectively, for each class. The training set is composed by 60% of the configurations and the testing set by 40% obtained from the JET database. The percentage of success is 100% for all the classes.

For the second classifier the number of magnetic configurations considered is 8. The class with a greater number of cases (VH_3M5_HT3) has been discarded to build a more

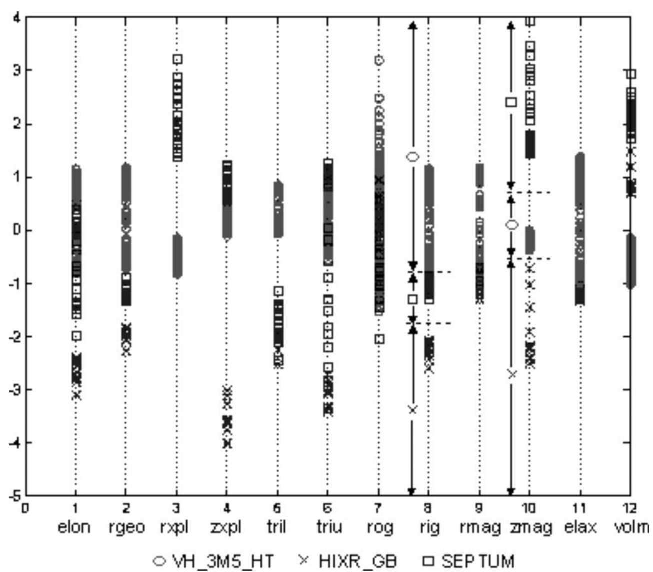

FIG. 1. Geometrical parameters of the analyzed data set for three plasma configurations.

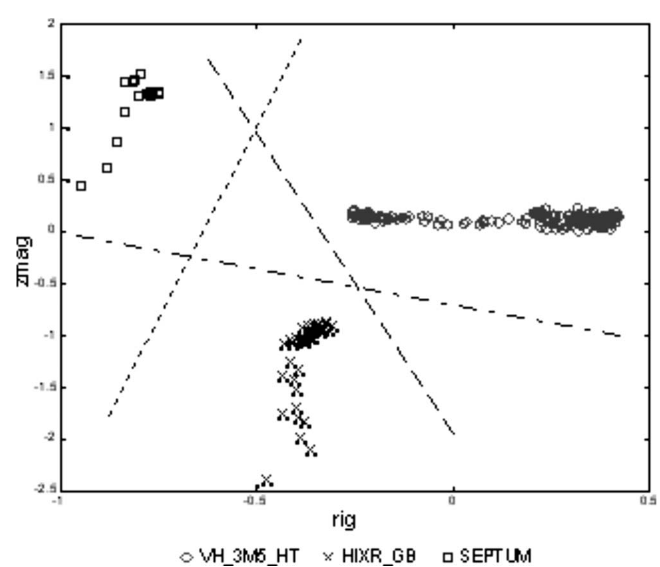

FIG. 2. Three classes described by two-dimensional feature vector (z_{mag} and r_{ig}).

Downloaded 11 Aug 2011 to 194.81.223.66. Redistribution subject to AIP license or copyright; see http://rsi.aip.org/about/rights_and_permissions

TABLE II. Results for the second classifier.

	Kernel		
	Linear	Radial basis $\sigma=100$	Exponential radial basis $\sigma=100$
HIXR_GB	96.6	95.3	96
StandardFat	100	100	100
HC_SFE_LT	94	95.5	92.2
SEPTUM	95	90	95
VLPC_SWEEP	100	100	100
D1F_C_SFE_LT	100	100	100
V_LFE_LT	98	96.2	98.7
D1Z_XFORM	87.5	90	87.5

uniform data set. The data set is composed of 39, 9, 24, 17, 45, 24, 40, and 12 configurations for each class. In this case, we use as feature vectors all the parameters described in Table I. The percentage of correct classifications is illustrated in Table II for three different kernels.

V. DISCUSSION

SVM is a very competitive method to classify JET configurations. Although apparently a simple visual inspection could be enough to discriminate a limited number of different discharges, when a bigger number of categories are considered, it is necessary to resort to a general purpose system as SVM. High success rate in spite of the reduced number of training data should be emphasized. Results with eight classes are promising even for a future real-time application of the method.

ACKNOWLEDGMENTS

This work, supported by the European Communities under the contract of Association between EURATOM/CIEMAT, was carried out within the framework of the European Fusion Development Agreement. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

[1] V. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed. (Springer, New York, 2000).
[2] R. Duda, P. Hart, and D. Store, *Pattern Classification*, 2nd ed. (Wiley, New York, 2001).
[3] L. Bottou, O. Chapelle, D. DeCoste, and J. Weston, *Large_Scale Kernel Machines* (MIT, Cambridge, MA, 2007).

Downloaded 11 Aug 2011 to 194.81.223.66. Redistribution subject to AIP license or copyright; see http://rsi.aip.org/about/rights_and_permissions