Fusion Engineering and Design 96–97 (2015) 907–911

Contents lists available at ScienceDirect

Fusion Engineering and Design

journal homepage: www.elsevier.com/locate/fusengdes

Feature selection for disruption prediction from scratch in JET by using genetic algorithms and probabilistic predictors

Augusto Pereira[a,*], Jesús Vega[a], Raúl Moreno[a], Sebastián Dormido-Canto[b], Giuseppe A. Rattá[a], Fernando Pavón[b], JET EFDA Contributors[1]

[a] Laboratorio Nacional de Fusión, CIEMAT, Madrid, Spain
[b] Dpto. Informática y Automática – UNED, Madrid, Spain

ARTICLE INFO

Article history:
Received 19 September 2014
Received in revised form 13 April 2015
Accepted 16 April 2015
Available online 23 May 2015

Keywords:
Feature selection
Genetic algorithm
Fitness function
Probabilistic predictor

ABSTRACT

Recently, a probabilistic classifier has been developed at JET to be used as predictor from scratch. It has been applied to a database of 1237 JET ITER-like wall (ILW) discharges (of which 201 disrupted) with good results: success rate of 94% and false alarm rate of 4.21%. A combinatorial analysis between 14 features to ensure the selection of the best ones to achieve good enough results in terms of success rate and false alarm rate was performed. All possible combinations with a number of features between 2 and 7 were tested and 9893 different predictors were analyzed. An important drawback in this analysis was the time required to compute the results that can be estimated in 1731 h (~2.4 months).

Genetic algorithms (GA) are searching algorithms that simulate the process of natural selection. In this article, the GA and the Venn predictors are combined with the objective not only of finding good enough features within the 14 available ones but also of reducing the computational time requirements.

Five different performance metrics as measures of the GA fitness function have been evaluated. The best metric was the measurement called Informedness, with just 6 generations (168 predictors at 29.4 h).

1. Introduction

A disruption is a dramatic extinction of the plasma current with a sudden loss of confinement in a very short time scale. It can produce large forces, strong loads and irreversible damage to the fusion device and surrounding components. Disruption avoidance is extremely important in Tokamak devices. In JET, machine learning methods have been used for the development of automatic systems able to predict incoming disruptions [1]. The objective of a disruption predictor is to predict as early as possible the plasma behavior like disruptive while a discharge is in execution. The Advance Predictor Of DISruptions (APODIS) was developed as a data-driven model based on a combination of several support vector machine (SVM) classifiers. It has been installed in the JET real-time data network [2] and has been very successful during the last ILW campaigns [3]. An optimized APODIS system to predict from scratch [4] obtained also good prediction rates. From scratch means that there is a lack of information during the first training processes and the predictor has to learn without any knowledge about disruptions from the beginning. It happens when a new machine starts its operation (great relevance thinking of ITER). Continuing with this paradigm, recently, a high learning rate probabilistic predictor (based on Venn machines) has been developed at JET [5] under the 'from scratch' approach. Venn predictors are used as probabilistic classifiers instead of using bare classifiers like SVM. They have the advantage of providing a confidence level for each individual prediction. Moreover, in the specific implementation carried out in [5], it has been possible to reduce the input sample space by means of grouping the data to use a nearest centroid taxonomy (NCT). To concentrate as much as possible the information of both plasma behaviors (the disruptive and non-disruptive activities), the Venn predictor uses the NCT that is fully described in Ref. [5]. This results in a faster predictor when testing new discharges. Success rate of 94% and false alarm rate of 4.21% were achieved from a database of 1237 discharges, of which 201 disrupted, belonging to ILW campaigns at JET. This implementation uses seven signals to characterize the disruptive/non-disruptive plasma state. These signals are processed using 32 ms time windows with a sampling frequency of 1 kHz. Two features per signal are computed during the 32 ms time windows: mean value and standard deviation of the

* Corresponding author. Tel.: +34 913460929.
E-mail address: augusto.pereira@ciemat.es (A. Pereira).
[1] See the Appendix of F. Romanelli et al., Proceedings of the 24th IAEA FEC 2012, San Diego, USA.

http://dx.doi.org/10.1016/j.fusengdes.2015.04.040
0920-3796/

Fourier spectrum (after removing the DC component). To ensure the selection of the best features between the 14 available ones, a combinatorial analysis was performed. All possible combinations with a number of features between 2 and 7 were tested and 9893 different predictors were analyzed. It was a costly process and the main drawback was the time required to calculate these results. The computational time was 1731 h (equivalent to 2.4 months). In this work and with the aim of reducing this time as much as possible, a feature selection method based on GA and the Venn predictors are combined. Moreover, different performance measures are evaluated, demonstrating that it is really significant the correct selection of the metric to get quick and successful results. The paper is organized in five sections. Section 2 explains several metrics to assess learning algorithms that can be used as fitness function on GA. Section 3 describes the genetic search procedure as an important tool to find impactful variables. Section 4 analyses the results of the five performance metrics used with GA and finally, Section 5 summarizes the most relevant contributions in the present paper.

2. Performance measure

A classifier is, typically, evaluated by a confusion matrix as illustrated in Fig. 1a. The columns are the actual class and the rows are the predicted class. TN is the number of negative examples correctly classified (also called true negatives or non-disruptive discharges in the present case). *FP* is the number of negative examples incorrectly classified as positive (false positives or false alarms), *FN* is the number of positive examples incorrectly classified as negative (false negatives or missed alarms) and *TP* is the number of positive examples correctly classified (true positives or disruptive discharges). Therefore, the predictive accuracy can be defined as follows.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

But, *Accuracy* might not be appropriate when test data are imbalanced and/or the costs of different errors vary markedly [6]. The receiver operating characteristic (ROC) curve is a standard technique for summarizing classifier performance over a range of tradeoffs between *TP* and *FN* error rate. On a ROC curve, the *x*-axis represents the *FP* rate or false alarm rate (*FA*), i.e. $FA = FP/(TN + FP) = 1 - TN/(TN + FP) = 1 - specificity$ and the *y*-axis represents the *recall* or success rate (*SR*), i.e. $SR = TP/(TP + FN)$, also called *sensitivity*. With a more clarify terminology, *SR* is the quotient of disruptive discharges correctly classified by the overall number of disruptive discharges that have been classified (the correctly and the incorrectly ones). *FA* is the same previous relationship but taking the non-disruptive discharges fails instead of the disruptive discharges hits. The ideal point on the ROC curve would be (0, 1), that is, all positive examples are classified correctly and no negative examples are misclassified as positive, $SR - FA = 1$, but in practice, in the most cases, it does not happen so. The precision for a class is the positive predictive value, i.e. $precision = TP/(TP + FP)$ and the main goal is to improve the *sensitivity* without hurting the *precision*, but this objective can be often conflicting, since when increasing the *TP* for the minority class, the *FP* for the majority class can also be increased (Fig. 1b); this will reduce the *precision*. The *F1-score* metric is one that combines the tradeoffs of *precision* and *sensitivity*.

$$F1\text{-}score = \frac{2TP}{2TP + FP + FN} \quad (2)$$

However, *Accuracy* and *F1-score* can present specific biases, namely that they ignore performance in correctly handling negative examples, i.e. $NPV = TN/(TN + FN)$, thus, *Informedness*, *Markedness* and *MCC* (Matthew's correlation coefficient) are formulated in [7] as unbiased measures to avoid the bias of *sensitivity*, *precision* and *Accuracy* measures, respectively.

$$Informedness = sensitivity + specificity - 1 \quad (3)$$

$$Markedness = precision + NPV - 1 \quad (4)$$

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (5)$$

Regardless of their suitability, the above five equations are contrasted in the fourth section as fitness function that can be used properly on GA.

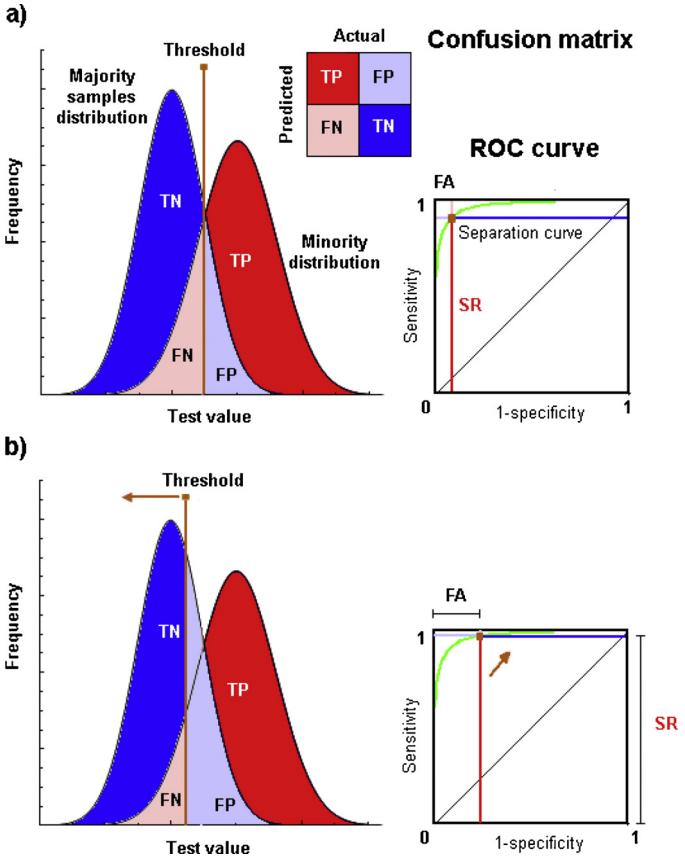
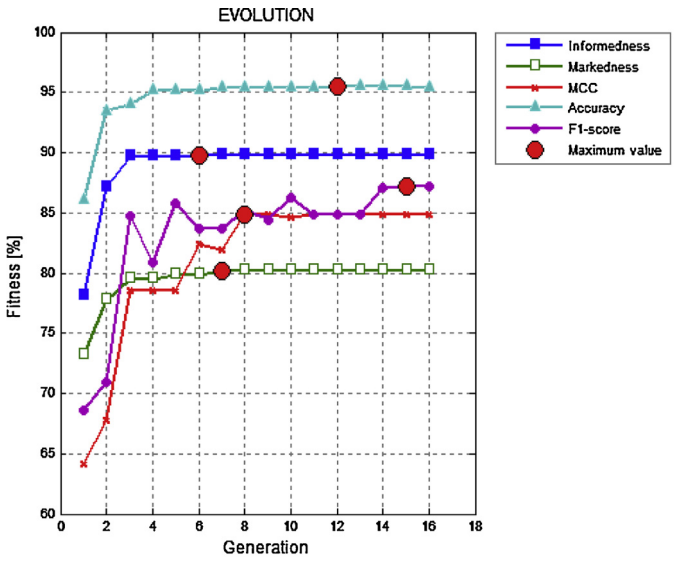


Fig. 1. Useful information to assess learning algorithms.

3. Feature selection and genetic algorithms

Feature selection, is the process of selecting the most important features, from a large set of them, by eliminating the redundant and irrelevant ones. Redundant features are those which provide no more information than the currently selected features, and irrelevant features provide no useful information in any context. Both of them may have negative effects on classification algorithms, increasing computational time and reducing accuracy and recognition rate. GA has demonstrated the effectiveness to identify variables in the data set as important [8], providing good results in limited computational time. On the other hand, to find the best solution implies testing the whole combination of features and it requires high computational costs consuming a lot of time. This paper presents a fast method using GA for feature extraction. At the present technique, the algorithm starts with an initial set of random solution called population. A random population of 28 individuals is generated. Each individual is also known as chromosome and it is formed by 14 genes (or features). The quality of each

chromosome is estimated by a classifier. A probabilistic classifier based on Venn predictors [5] is used. Venn predictors make prediction directly from data by transduction (without generating any rules), instead of repeatedly training classifiers to generate models. Previously, the input sample space (all disruptive/non-disruptive information) is condensed in a proper way (NCT), resulting thus, in a faster and balanced classifier for testing the whole population. At each generation (iteration) of the GA, the 28 chromosomes are evaluated using the fitness function. It plays the most important role in genetic search since offspring for the next generation are determined by the fitness function score, which reflects how optimal the solution is: the higher the number the better the solution. This function has to assess the goodness of the numerical error rates obtained with the classifier. Thus, the output of the classifier is the input of the fitness function and it returns a numerical evaluation representing the goodness of the feature subset. In this paper we want to emphasize the importance of choosing a metric to use as fitness function. The five performance metrics explained in the previous section were used as measures of the GA fitness function. After that, the best individuals with the best scores are selected for creating the next generation. Two more genetic operators such as crossover and mutation will explore new regions of search space by the combination and the replacing of genes, while keeping some of the current information at the same time (with 5% mutation probability and 55% crossover probability). The generational process ends when the termination criterion is satisfied, for instance, the number of generations (just 50 times). Finally, the winner features are the members of those individuals who reach the highest fitness function, from the first to the last generation. Indeed, more than one individual can have the same maximum fitness value (various subsets of features can be winners with the same score).



METRIC	EVALUATED PREDICTORS	ELAPSED TIME (hours)	GENERATIONS
Informedness	168	29.4	6
Markedness	196	34.3	7
MCC	224	39.2	8
Accuracy	336	58.8	12
F1-score	420	73.5	15

Fig. 2. GA evolution using five different metrics. Bigger dots represent the first generation with the maximum fitness value on every evolution.

4. Evaluation and results

The list of signals is explained in Table 1, which is the same as used in the work [5]. As previously, it consists of 14 features belongs to 7 plasma signals.

For each signal, is included the standard deviation of the Fourier spectrum after removing the DC component. As demonstrated and explained in [1,3,5], the use of frequency domain with that procedure enhances the ability of prediction. According to Table 2, each x corresponds to a signal being turned on. The best candidates for predictions are the ones with *FA* of 4.21%. Among the five candidates, the predictors with lower number of features are preferred because it will take less time to try new observations and therefore a better fault tolerant. This reduces the possibilities to only two options. From a physical point of view, ML and I_p signals are essential components (features 2, 3 and 4). But, they still need either LI or Pout signal to warn for disruptions in a more relevant way. The final selection between them is favorable to the predictor with the smaller probability error bar (the predictor with features 2, 3, 4 and 5) because it means a more accurate prediction. Nevertheless, a genetic search using a data-driven model must be able to find the five candidates with the best score reached in terms of success rate and false alarm rate with values of 94.00% and 4.21% respectively. At the present work, a random population of 28 individuals (twice the number of features) is generated. By using the same initial population, the genetic search is performed five times, once for each of the five different fitness functions. Prediction from scratch of a single individual with the whole dataset (1237 discharges) lasts 10.5 min; therefore a single generation takes about 4.9 h.

The evaluation of the five metrics using the combination of GA and Venn predictors is presented in Fig. 2. Fitness (%) is the value of each metric in percentage. *Informedness* metric matched the highest fitness value with just six generations (168 predictors were assessed in 29.4 h); on the other hand, the choice of *F1-score* needed fifteen generations, equivalent to assess 420 predictors in 73.5 h. Table 3 displays the best features by using *Informedness* measurement as fitness function. It can be appreciated how the first best rate appears on the sixth generation and the last one, with the same maximum rate but different subset of features, arises on the 19th generation (532 predictors at 93.1 h), see Fig. 3.

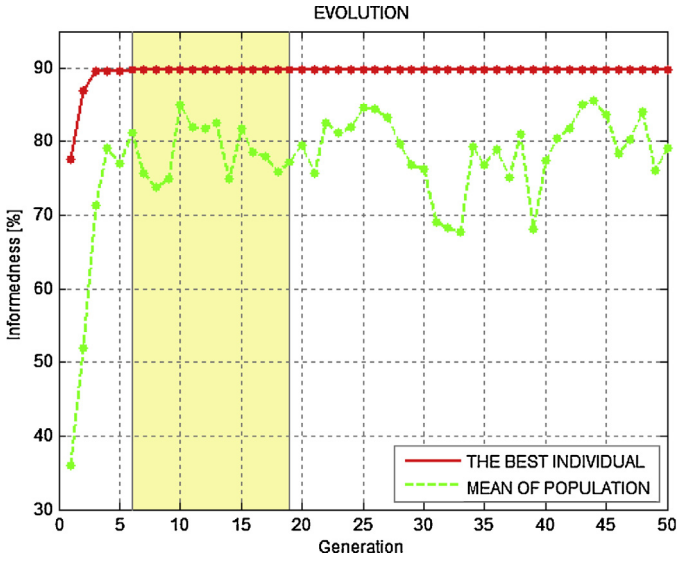


Fig. 3. Highlight region denotes where first emerged all individuals with the most relevant features.

Table 1
List of signals and features.

Signal name	Label	Units	Id	Definition
Plasma current	Ip	A	1	mean(Ip)
			2	std(|fft(Ip)|)
Mode locked amplitude	ML	T	3	mean(ML)
			4	std(|fft(ML)|)
Plasma internal inductance	LI		5	mean(LI)
			6	std(|fft(LI)|)
Plasma density	Ne	m^{-3}	7	mean(Ne)
			8	std(|fft(Ne)|)
Diamagnetic energy derivative	dW/dt	W	9	mean(dW/dt)
			10	std(|fft(dW/dt)|)
Radiated power	Pout	W	11	mean(Pout)
			12	std(|fft(Pout)|)
Total input power	Pin	W	13	mean(Pin)
			14	std(|fft(Pin)|)

Table 2
Features with the best scores reached in a previous work [5]. The average prediction probability (AVP) represents the average probability interval in Venn predictors.

Feature Id														*SR* (%)	*FA* (%)	AVP
1	2	3	4	5	6	7	8	9	10	11	12	13	14			
		x	x											94.00	4.70	0.813 ± 0.187
	x		x											92.50	5.09	0.831 ± 0.169
	x	x	x											94.00	4.31	0.809 ± 0.191
		x	x							x				94.00	4.70	0.813 ± 0.187
	x	x	x	x										**94.00**	**4.21**	0.811 ± 0.189
	x	x	x							x				**94.00**	**4.21**	0.810 ± 0.190
	x	x	x	x						x				**94.00**	**4.21**	0.811 ± 0.189
	x	x	x			x	x							**94.00**	**4.21**	0.803 ± 0.197
	x	x	x			x	x			x				**94.00**	**4.21**	0.803 ± 0.197
	x	x	x	x						x	x			94.00	4.31	0.810 ± 0.190
	x	x	x	x		x	x			x				94.00	4.31	0.802 ± 0.198
	x	x	x			x	x			x	x			94.00	4.31	0.802 ± 0.198

Bold: the highest rates.

Table 3
Features sorted by Informedness.

Informedness (%)	*SR* (%)	*FA* (%)	Features														Generation (%)
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	
89.79	**94.00**	**4.21**	0	1	1	1	0	0	1	1	0	0	1	0	0	0	**6**
89.79	**94.00**	**4.21**	0	1	1	1	0	0	1	1	0	0	0	0	0	0	9
89.79	**94.00**	**4.21**	0	1	1	1	1	0	0	0	0	0	1	0	0	0	13
89.79	**94.00**	**4.21**	0	1	1	1	0	0	0	0	0	0	1	0	0	0	18
89.79	**94.00**	**4.21**	0	1	1	1	1	0	0	0	0	0	0	0	0	0	**19**
89.69	94.00	4.31	0	1	1	1	0	0	0	0	0	0	1	1	0	0	6
89.69	94.00	4.31	0	1	1	1	0	0	0	1	0	0	0	0	0	0	7
89.69	94.00	4.31	0	1	1	1	0	0	1	1	0	0	0	1	0	0	10
89.69	94.00	4.31	0	1	1	1	0	0	1	1	0	0	1	1	0	0	11
89.69	94.00	4.31	0	1	1	1	1	0	1	1	0	0	1	0	0	0	11
89.69	94.00	4.31	0	1	1	1	0	0	1	0	0	0	1	0	0	0	14
89.69	94.00	4.31	0	1	1	1	1	0	1	0	0	0	1	0	0	0	15
89.69	94.00	4.31	0	1	1	1	0	0	0	1	0	0	1	0	0	0	15
89.69	94.00	4.31	0	1	1	1	1	0	1	1	0	0	0	0	0	0	19
89.69	94.00	4.31	0	1	1	1	0	0	0	0	0	0	0	0	0	0	19
89.69	94.00	4.31	0	1	1	1	1	0	0	0	0	0	1	1	0	0	25
89.69	94.00	4.31	0	1	1	1	1	0	0	1	0	0	1	0	0	0	27
89.69	94.00	4.31	0	1	1	1	0	0	1	0	0	0	0	0	0	0	29
89.69	94.00	4.31	0	1	1	1	1	0	1	0	0	0	0	0	0	0	35
89.69	94.00	4.31	0	1	1	1	1	0	0	1	0	0	0	0	0	0	47

Bold: the highest rates.

5. Conclusions

Five performance measures were evaluated as GA fitness function.

The most relevant characteristics obtained in this analysis are consistent with those achieved previously [5], but with the difference of significantly reducing the employed time (1731 vs. 29.4 h), an improvement of 98.31%. *Informedness*, *Markedness* and *MCC* show better performance than *Accuracy* and *F1-score* metrics, finding in less time the most impactful variables to correctly use on disruption prediction. The first three ones (cataloged as unbiased measurements) are more objectives handling incorrectly classified examples and the measures are better weighted. Therefore, it is really significant the correct selection of the fitness function on GA to get quick and successful results.

Acknowledgments

This work was partially funded by the Spanish Ministry of Economy and Competitiveness under the Projects No. ENE2012-38970-C04-01 and ENE2012-38970-C04-03.

This work was supported by EURATOM and 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.

References

[1] G.A. Rattá, J. Vega, A. Murari, G. Vagliasindi, M.F. Johnson, P.C. de Vries, et al., An advanced disruption predictor for JET tested in a simulated real time environment, Nucl. Fusion 50 (2010) 025005.

[2] J.M. López, J. Vega, D. Alves, S. Dormido-Canto, A. Murari, J.M. Ramirez, et al., Implementation of the disruption predictor APODIS in JET real-time network using the MARTe framework, in: 18th Real-Time Conference, 11–15 June 2012, Berkeley, CA, USA, 2012, http://rt2012.lbl.gov/RT2012Abstracts.pdf. Submitted to IEEE Trans. on Nuclear Science.

[3] J. Vega, S. Dormido-Canto, J.M. López, A. Murari, J.M. Ramírez, R. Moreno, et al., Results of the JET real-time disruption predictor in the ITER-like wall campaigns, Fusion Eng. Des. 88 (6–8) (2013) 1228–1231.

[4] S. Dormido-Canto, J. Vega, J.M. Ramírez, A. Murari, R. Moreno, J.M. López, et al., Development of an efficient real-time disruption predictor from scratch on JET and implications for ITER, Nucl. Fusion 53 (2013) 113001.

[5] J. Vega, A. Murari, S. Dormido-Canto, R. Moreno, A. Pereira, A. Acero, et al., Adaptive high learning rate probabilistic disruption predictors from scratch for the next generation of tokamaks, Nucl. Fusion 54 (12) (2015) 123001.

[6] N.V. Chawla, Data Mining for Imbalanced Datasets: An Overview. Data Mining and Knowledge Discovery Handbook, 2005, ISBN 978-0-387-24435-8, pp. 853–867.

[7] M.W. David, Evaluation: from precision, recall and F-measure to ROC, Informedness, Markedness & Correlation, J. Mach. Learn. Technol. 2 (1) (2011) 37–63, ISSN: 2229-3981.

[8] G.A. Rattá, J. Vega, A. Murari, JET EFDA contributors, Improved feature selection based on genetic algorithms for real time disruption prediction on JET, Fusion Eng. Des. 87 (September (9)) (2012) 1670–1678.