Assessment of linear disruption predictors using JT-60U data

J. Vega[a], F. Hernández[b], S. Dormido-Canto[b], A. Isayama[c], E. Joffrin[d], G. Matsunaga[c], T. Suzuki[c]

[a]*Laboratorio Nacional de Fusión – CIEMAT. 28040 Madrid, Spain*
[b]*Depto. De Informática y Automática. UNED. 28040 Madrid, Spain*
[c]*QST Fusion Energy Research and Development Directorate. Naka, Japan*
[d]*CEA, IRFM, F-13108 St Paul Les Durance, France*

Disruptions are dangerous events in tokamaks that require mitigation methods to alleviate its detrimental effects. A prerequisite to trigger any mitigation action is the existence of a reliable disruption predictor. This article assesses a predictor that relates in a linear way consecutive samples of a single quantity (in particular, the magnetic perturbation time derivative signal has been used). With this kind of predictor, the recognition of disruptions does not depend on how large the signal amplitude is but on how large the signal increments are: small increments mean smooth plasma evolution whereas abrupt increments reflect a non-smooth evolution and potential risk of disruption. Results are presented with data from the JT-60U tokamak and high-beta discharges. Two training methods have been tested: a classical approach in which the more data for training the better and an adaptive method that starts from scratch. In both cases the success rate is about 95%. It should be noted that predictors based on signal increments and their adaptive versions can be of big interest for next devices such as JT-60SA or ITER.

Keywords: disruption prediction, signal increments, nearest centroid, JT-60SA, ITER, nuclear fusion.

1. Introduction

So far, the mode lock (ML) signal amplitude is one of the simplest disruption predictors used in tokamaks, where the rise of this amplitude above a certain level can be considered a precursor of a forthcoming disruption. Rotating magnetohydrodynamic (MHD) modes can be decelerated and become locked to the wall. The ML signal increases when the MHD mode rotation decreases or when the MHD mode amplitude grows. Therefore, an alarm is triggered when the ML amplitude is above a certain threshold (that can be set-up prior to each discharge). However, this simple predictor is not able to cope with success rates close to 100%.

An alternative to ML thresholds is the development of data-driven models by means of general machine learning methods (for example neural networks, Support Vector Machines (SVM) and deep learning among others). Machine learning methods split the parameter space into two zones (disruptive/non-disruptive) and determine the separation frontier between both zones. During the training process, the mathematical form of the frontier is obtained. During the execution of a discharge, points are projected into the parameter space and an alarm is raised when the point appears in the disruptive zone.

The main drawback of disruption predictors based on data-driven models resides on the fact that the mathematical form of the separating frontier is quite complex [1] and it does not allow a physics interpretation of the disruption. In other words, it is possible to predict disruptions but the physics reasons remain unknown [2].

The objective of this article is to test two different approaches of a new kind of disruption predictors that are not based on either general machine learning methods or amplitude thresholds. The rationale for the new type of predictors has been to follow the experimental evidence provided by magnetic signals close to a disruption. For example, the mode lock or the magnetic perturbation time derivative signals show abrupt changes in their amplitudes between consecutive samples and, therefore, large signal variations can be used as disruption precursors.

Section 2 describes the mathematical formulation of the predictor and section 3 explains the training procedure. Section 4 provides the assessment of the predictor under two different approaches and section 5 is a short discussion.

2. Mathematical formulation of the predictor

The predictor has been tested with a database of high beta experiments in the JT-60U tokamak, where the beta value was close or above the no-wall beta limit [3].

Specifically, as it is shown below, the predictor under test is based on signal increments and it uses the magnetic perturbation time derivative (MPTD) signal with a sampling period $\tau = 1$ ms. The magnetic signal is obtained by 8 saddle loops, which are located at mid-plane on the outboard side and distributed almost at regular intervals toroidally. The signal used for the analysis is the amplitude of the n=0 component of the saddle loop signal, where n is the toroidal mode number. Thus the MPTD amplitude is thought to be a good indicator of a sudden change of the plasma like a disruption. Specifications of the JT-60U magnetic sensors can be found in http://www-jt60.naka.qst.go.jp/english/diag/html/diag_56.html and in [4].

author's email: jesus.vega@ciemat.es

Fig. 1 shows the plasma current and the MPTD signal of a non-disruptive discharge. Because the aim is to recognize abrupt changes between consecutive samples, the parameter space of the predictor is a 2-dimensional space where the y coordinate of each point is the $MPTD(t)$ amplitude at a certain time t and the x coordinate is the $MPTD(t-\tau)$ amplitude of the previous sample, where τ is the sampling period. Fig. 1c is the parameter space corresponding to the whole discharge. It should be noted that the MPTD signal shows oscillations (fig. 1b) but the differences between consecutive samples are small and all the points in the scatterplot are located around the point (0, 0).

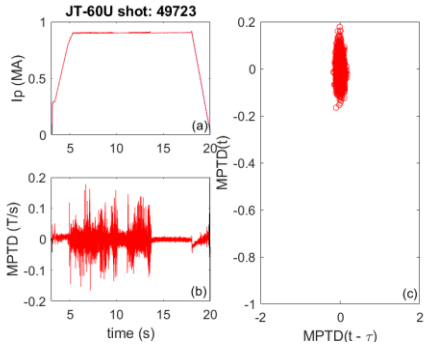

Fig. 1. (a) Plasma current of a non-disruptive discharge. (b) Magnetic perturbation time derivative. (c) Predictor parameter space.

Fig. 2 shows a disruptive discharge. Points in the parameter space are concentrated around (0, 0), but they appear far from the origin when the disruption approaches.

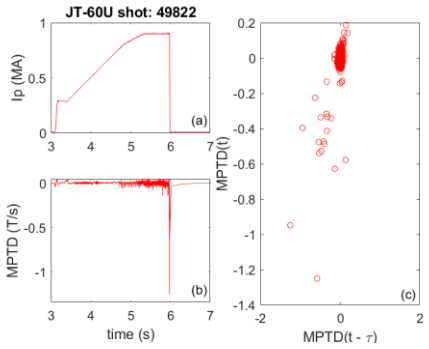

Fig. 2. (a) Plasma current of a disruptive discharge. (b) Magnetic perturbation time derivative. (c) Predictor parameter space.

Figures 1 and 2 suggest the possibility of condensing the disruptive and non-disruptive character of the discharges into two single points (or centroids) in the parameter space. The non-disruptive centroid will be located close to (0, 0) but the disruptive centroid will appear far from (0, 0).

To compress the disruptive and non-disruptive behaviors into two centroids, a training dataset from past discharges has to be used. After computing the centroids, points can be projected in real-time to the plane $MPTD(t-\tau)-MPTD(t)$ during the execution of discharges. As the sampling period is τ, there will be a new projection $P(x_1, x_2)$ in the parameter space with this period (fig. 3). The predictor output will be '*disruptive*' ('*non-disruptive*') when the nearest centroid to P is the disruptive (non-disruptive) centroid.

Therefore, given the coordinates of the disruptive $C_D(d_1, d_2)$ and non-disruptive $C_N(c_1, c_2)$ centroids together with the projected point $P(x_1, x_2)$ at time t, $x_1 = MPTD(t-\tau)$ and $x_2 = MPTD(t)$, the plasma behavior is disruptive when

$$E_{P,C_D} < E_{P,C_N}$$

where $E_{P,C}$ are the Euclidean distances. In other words, the condition of disruptive behavior is given by

$$\sqrt{(x_1-d_1)^2+(x_2-d_2)^2} < \sqrt{(x_1-c_1)^2+(x_2-c_2)^2}$$

After simple algebraic manipulations

$$x_2 > -\frac{d_1-c_1}{d_2-c_2}x_1 + \frac{d_1^2+d_2^2-c_1^2-c_2^2}{2(d_2-c_2)} \qquad (1)$$

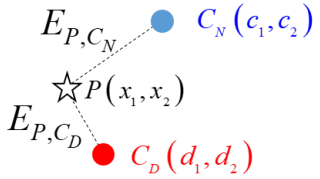

Fig. 3. The point P is a projection at time t in the parameter space. If P is closer to the disruptive centroid, it identifies a disruptive behavior. Otherwise, the plasma evolves in a non-disruptive way.

Equation (1) establishes a linear relationship between the amplitudes x_2 and x_1 to identify an incoming disruption and it has been obtained according to the nearest centroid approach.

3. Determination of centroids

The computation of centroids has to be considered as a training process. Also, it is important to note that the non-disruptive centroid and the disruptive one are computed in a different way.

3.1 Non-disruptive centroid

Given a dataset of N_N non-disruptive discharges, one centroid is computed per non-disruptive shot. To this end, all pairs

$$\{(MPTD(t_1), MPTD(t_1+\tau)),$$
$$(MPTD(t_1+2\tau), MPTD(t_1+3\tau)),$$
$$\ldots\ldots,$$
$$(MPTD(t_n-\tau), MPTD(t_n))\}$$

of each discharge between t_1 and t_n are considered, where t_1 is the first time in which the plasma current is above 0.1 MA and t_n is the last time in which the plasma

current is above 0.1 MA (it should be noted that an even number of samples are required and, perhaps, the y coordinate of the last pair could be the last but one sample with plasma current above 0.1 MA). Therefore, N_N individual centroids are determined, $C_j\left(C_{j,X}, C_{j,Y}\right), j = 1,..., N_N$, where

$$C_{j,X} = mean\left\{MPTD\left(t_1 + 2K\tau\right)\right\}, K = 0,1,2,... ,$$

$$C_{j,Y} = mean\left\{MPTD\left(t_1 + \left(2K+1\right)\tau\right)\right\}, K = 0,1,2,...$$

and $mean\{\bullet\}$ is the mean value. Once that the N_N individual centroids have been determined, the global centroid that condense all the non-disruptive behavior is

$C_N\left(c_1 = mean\left\{C_{j,X}\right\}, c_2 = mean\left\{C_{j,Y}\right\}\right), j = 1,..., N_N$. According to the reasoning of section 2 regarding non-disruptive discharges, $\left(c_1, c_2\right) \approx \left(0,0\right)$.

3.2 Disruptive centroid

Given a dataset of N_D disruptive discharges, one centroid per discharge is determined in the parameter space: $C_j\left(d_{j,X}, d_{j,Y}\right), j = 1,..., N_D$. The coordinate $d_{j,Y}$ is the minimum amplitude of the MPTD signal within a time window 20 ms long around the disruption time (fig. 4). The coordinate $d_{j,X}$ is the previous sample to $d_{j,Y}$.

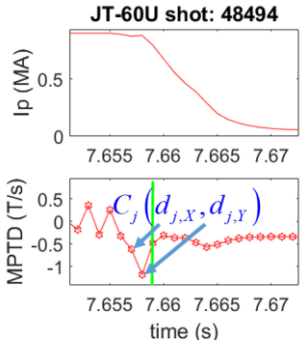

Fig. 4. Centroid of a disruptive discharge. The green line shows the disruption time.

The global disruptive centroid C_D is the mean value of the individual centroids, *i.e.*

$$C_D\left(d_1 = mean\left\{d_{j,X}\right\}, d_2 = mean\left\{d_{j,Y}\right\}\right), j = 1,..., N_D$$

4. Predictor assessment

A dataset of 154 discharges (76 unintentional disruptions and 78 non-disruptive shots) in the range 47756 – 49826 and corresponding to high beta operation of JT-60U has been used. The linear predictor has been tested under two different training conditions. The first one utilizes approximately 40% of disruptive and non-disruptive discharges for training and the rest for test. The second training condition uses an adaptive training as explained in section 4.2. In all cases, the predictor triggers an alarm when two consecutive predictions recognize a disruptive behavior. This has been established to reduce the number of false alarms without affecting the success rate.

Five parameters have been considered to qualify the tests:

GSR (%): global success rate: number of disruptions identified over the total number of disruptions in the test set.

SRP (%): success rate with positive warning time: number of disruptions identified before the disruption over the total number of disruptions in the test set.

MA (%): missed alarm rate: number of disruptive discharges not recognized as disruptions over the total number of disruptive shots.

FA (%): false alarm rate: number of non-disruptive discharges that triggered an alarm over the total number of non-disruptive discharges.

AVGWT (ms): average warning time: anticipation time (on average) of the alarms before the disruption.

4.1 Non-adaptive approach

The training and test datasets have been chosen at random from the available set of discharges. In particular, 32 disruptive and 30 non-disruptive discharges have been used for training purposes (*i.e.* to compute the centroids) and 44 disruptive and 48 non-disruptive discharges for test.

Fig. 5 shows the individual centroids of the training discharges (circles) together with the global centroids (crosses). It is important to note that the non-disruptive centroids are located practically in the same point.

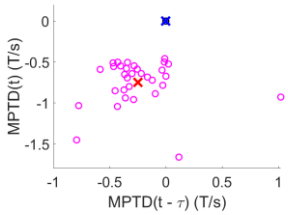

Fig. 5. The blue and red crosses are, respectively, the global centroids of non-disruptive and disruptive shots.

Row 1 of table 1 shows the results obtained with the previous dataset of discharges. However, to remove any possible bias in the random selection of shots, 10 additional random sets for training/test were chosen. The outcomes appear from row 2 to 10. The results are quite similar in all cases and the global result of the linear predictor are shown in the last row.

Table 1. Prediction results with 11 random selections of training/test datasets.

#test	GSR	SRP	MA	FA	AVGWT
1	100	97.7	0	6.2	20
2	97.7	97.7	2.3	4.2	17
3	100	97.7	0	4.2	13
4	95.5	90.9	4.5	4.2	14
5	97.7	95.5	2.3	2.1	14
6	97.7	90.9	2.3	8.3	19
7	100	97.7	0	6.2	18

8	97.7	95.5	2.3	10.4	20
9	100	97.7	0	8.3	18
10	95.5	90.9	4.5	0	17
11	100	97.7	0	4.2	20
Mean	**98.3**	**95.4**	**1.7**	**5.3**	**17**

With the centroids of fig. 5, equation (1) is:

$$x_2 < -0.3300 \cdot x_1 - 0.4152$$

where $x_2 = MPTD(t)$ and $x_1 = MPTD(t-\tau)$. Fig. 6 shows the parameter space of three different situations: a non-disruptive discharge, a successful recognition of a disruptive behavior and a false alarm. It is important to emphasize that most of the false alarms correspond to minor disruptions. In these cases, the plasma is able to recover but the predictor triggers an alarm because at least two consecutive points appear in the disruptive zone of the parameter space.

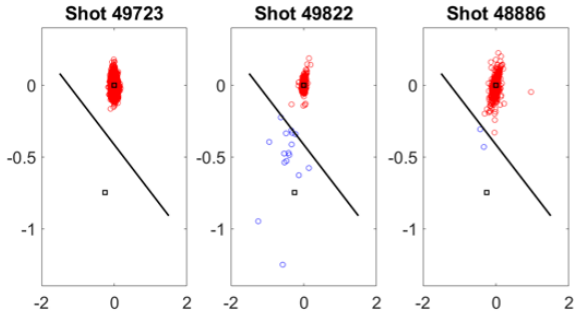

Fig. 6. Parameter spaces corresponding to a non-disruptive shot (left), disruptive shot (middle) and false alarm (right). Black squares are the centroids.

Fig. 7 plots the cumulative fraction of detected disruptions for the case of the first row of table 1. 100% of disruptions are recognized but one of them is identified 2 ms after the disruption (tardy detection). The cumulative fraction increases mainly for warning times less than 14 ms which makes this predictor suitable for mitigation purposes.

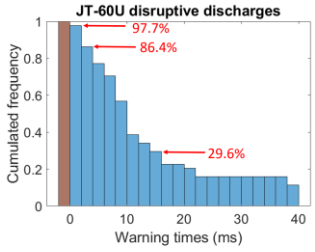

Fig. 7. Blue bins represent positive warning times and the brown one is a tardy detection. Bin width is 2 ms. 29.6%, 86.4% and 97.7% of the disruptions are recognized with a warning time greater than 14 ms, 2 ms and 0 respectively.

4.2 Adaptive approach from scratch

So far, there is neither theory from first principles nor a single physics model to explain the disruptive instability [5]. Therefore, each tokamak develops its own predictors from its own database of past discharges. An attempt to create cross-predictors were made in the past, for example between JET and AUG [6], but more work is necessary. Therefore, due to the lack of databases in new devices such as JT-60SA or ITER, predictors have to be created from scratch and have to learn in an adaptive way. Examples of disruption predictors from scratch can be found in references [7-9], all of them applied to JET data.

The objective of prediction from scratch is to process the discharges in chronological order, as it happens in real operation, and to create the first predictor after having (at least) 1 disruptive discharge and 1 non-disruptive discharge. The criterion used in this article to retrain the predictor establishes that it has to be carried out after each missed alarm.

The application of centroids to predict from scratch to the dataset of 154 discharges from JT-60U provides the results of table 2. The first predictor uses 1 disruptive discharge and 2 non-disruptive discharges. Only 2 disruptions are missed and, therefore, only 2 re-trainings have been necessary. The inequality from equation (1) to recognize disruptive behaviors is:

$$x_2 < -0.5580 \cdot x_1 - 0.3664$$

The adaptive approach seems to be very sensible to disruptive behaviors and this is the reason of the high rate of false alarms and the high value of the AVGWT parameter in comparison with the values of table 1. Possibly, this fact is a consequence of using discharges in a wide range of +2000 shots (154 discharges between 47756 and 49826) instead of learning with shots in a short range.

Table 2. Prediction results with adaptive prediction from scratch.

GSR	SRP	MA	FA	AVGWT
97.3	94.7	2.7	19.7	59

5. Discussion

Disruption predictors based on signal increments allow linear relationships to make simple predictions and to ensure straightforward real-time implementations. Two training approaches have been tested (non-adaptive and adaptive predictors) and the success rates are quite high in both approaches. In the case of adaptive predictions, more research is necessary to reduce the false alarm rate. This was also observed in [7].

Acknowledgments

This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training programme 2014-2018 under grant agreement No 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

This work was partially funded by the Spanish Ministry of Economy and Competitiveness under the Projects No. ENE2015-64914-C3-1-R and ENE2015-64914-C3-2-R.

References

[1] J. Vega, S. Dormido-Canto, J. M. López, A. Murari, J. M. Ramírez, R. Moreno et al. Fusion Engineering and Design 88 (2013) 1228-1231.
[2] Y. Zhang, G. Pautasso, O. Kardaun, G. Tardini, X. D. Zhang. Nucl. Fusion 51 (2011) 063039 (12pp).
[3] G. Matsunaga, K. Shinohara, N. Aiba, Y. Sakamoto, A. Isayama, N. Asakura et al. Nucl. Fusion 50 (2010) 084003 (8pp).
[4] Y. Neyatani, et al., "Development of magnetic sensors in JT-60 upgrade ", 14th Symp. on Fusion Eng., Vol.2 1183 (1991).
[5] A. H. Boozer, Phys. Plasmas 19, 058101 (2012).
[6] C. G. Windsor, G. Pautasso, C. Tichmann, R.J. Buttery, T. C. Hender. Nuclear Fusion 45 (2005) 337–350
[7] J. Vega, A. Murari, S. Dormido-Canto, R. Moreno, A. Pereira, A. Acero. Nuclear Fusion. 54 (2014) 123001 (17pp)
[8] S. Dormido-Canto, J. Vega, J. M. Ramírez, A. Murari, R. Moreno, J. M. López et al. Nuclear Fusion. 53 (2013) 113001 (8pp).
[9] A. Murari, M. Lungaroni, E. Peluso , P. Gaudio, J. Vega, S. Dormido-Canto et al. Nuclear Fusion 58 (2018) 056002 (16pp).