

## Automatic recognition of plasma relevant events: implications for ITER

Wednesday, May 15, 2019 5:05 PM (20 minutes)

Big data analytics deals with heterogeneous, complex and massive datasets to identify patterns that are hidden inside enormous volumes of data. ITER is expected to acquire more than 1 Tbyte of data per discharge. This amount of data comes from hundreds of thousands of signals acquired in each discharge. Signals can be time/amplitude series, temporal evolution of profiles and video-movies (infra-red and visible cameras). Therefore, the ITER database satisfies the conditions of heterogeneity, complexity, size and hidden patterns to use big data techniques.

ITER is a device not focused on basic research on plasma physics. Its aim is to produce high performance plasmas to approach the operation to reactor regimes. Vast amounts of hidden information will remain in the ITER databases and it will be worth to extract as much knowledge as possible from the data. Due to the large number of signals per discharge and the shot duration (30 minutes), automatic methods of data analysis will be necessary.

This work puts the focus on the use of big data algorithms for the automatic recognition of plasma relevant events in huge databases of nuclear fusion devices. A relevant event can be any kind of anomaly (or perturbation) in the plasma evolution. This is revealed in the temporal evolution signals as (more or less) abrupt variations (for instance in amplitude, noise, or sudden presence/suppression of patterns with periodical structure). Examples of events can be the input of additional power, gas injection, confinement transitions or diagnostic perturbative methods. Obviously, the automatic searching process for relevant events will have to find the above examples but the most interesting cases will be those ones whose temporal location is not explicitly related to known phenomena.

A first step to perform automatic data analysis is to recognise anomalies in individual signals. A second step is to determine characteristic times of anomalies. A third step is to identify the set of signals that show interesting patterns in the same intervals. As mentioned, the phenomenology behind these patterns can be unknown and it will be necessary to put together signals and time intervals that can correspond to the same type of physics event. To accomplish this, the fourth step is to use unsupervised clustering techniques to assign labels to each class of physics event. Once the clusters are formed, the fifth step is the creation of supervised learning classifiers for the automatic recognition of physics events. After having determined the different groups, there will be available sets of common signals and temporal locations where the same patterns appear. Therefore, there will be statistical relevance to look for a physics interpretation of the different labels found. It should be emphasised that after an off-line analysis, steps 1, 2 and the classifications resulting from step 5 can be carried out in real-time.

Specific details about these 5 steps will be given with emphasis on techniques of steps 1 and 2.

**Primary authors:** Dr VEGA, Jesús (Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas); CASTRO ROJO, Rodrigo (Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas); Dr DORMIDO-CANTO, Sebastián (UNED); Dr RATTÁ, Giuseppe A. (CIEMAT); RUIZ, Mariano (Universidad Politécnica de Madrid)

**Presenter:** Dr VEGA, Jesús (Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas)

**Session Classification:** Poster facult.

**Track Classification:** Advanced Computing and Massive Data Analysis