GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
Ciemat
Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas

Smart decimation method for fusion research data

R. Castro[1], J. Vega[1]

1) CIEMAT Fusion program. Avda. Complutense 40. Madrid. Spain

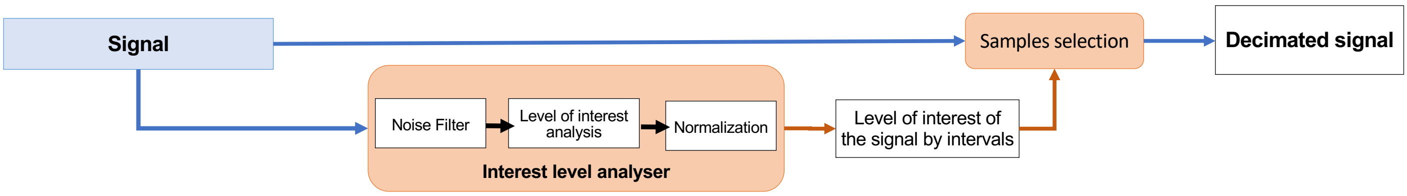
Abstract

New fusion research experiments will generate massive experimental data. For example, ITER will have above one million of variables from control signals and diagnostic systems. Some of these variables will produce data during long pulse (about 30 minutes) experiments and other will generate data continuously. Just to have a clearer idea of the problem, ITER estimates more than 10 GBytes/second of data flow during experiment pulses. In this context, fusion research appears in the scope of the big data where both search and access functions require new approaches and optimizations.

One common data access functionality is decimation. It allows to retrieve a limited number of points of the total. One classical mechanism of decimation is downsampling by an integer factor called step. It consists on selecting a value every 'step' number of values. The main characteristic of classical decimation is that selected values are uniformly distributed along the total. However, in case of time evolution experiment signals, the relevancy of data is not uniform. There are some intervals where the provided information is more complex and richer, and usually more interesting from user point of view.

The contribution presents a new data decimation technology for unidimensional time evolution signals where the limited number of accessed points are distributed based on data interest level. The new method implements, on one hand, a heuristic function which is able to determine the level of interest of an interval based on its data characteristics, and on the other hand, a selection algorithm where points are distributed based on weighted intervals. This level of interest analysis technique is part of a more global signal anomalies detection framework [1].

Smart Decimation method



Noise Filter

Function: To reduce original signal noise level and preserve as mush as possible original signal shape.

Method: Exponential Moving Averages [2]

Implementation: Iterative method

$$S(t)=\begin{cases} Y(0) & t=0 \\ \alpha Y(t)+(1-\alpha)S(t-1) & t>0 \end{cases}$$

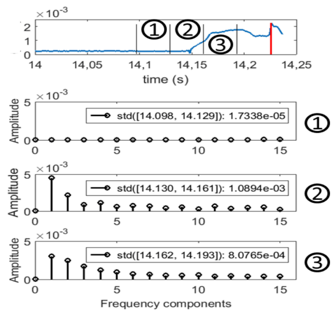
Level of interest analysis

Function: To measure the level of interest of different intervals of a signal.

Method: Standard deviation of the Fourier spectrum (positive frequencies) after removing the DC component. This technique

$$\mathbf{x}=(x_1,\ldots,x_n)\in \sim^n$$

$$x_i = std\left(\left|fft\left(s_i(t)\right)\right|_+^*\right), \quad i=1,\ldots,n$$



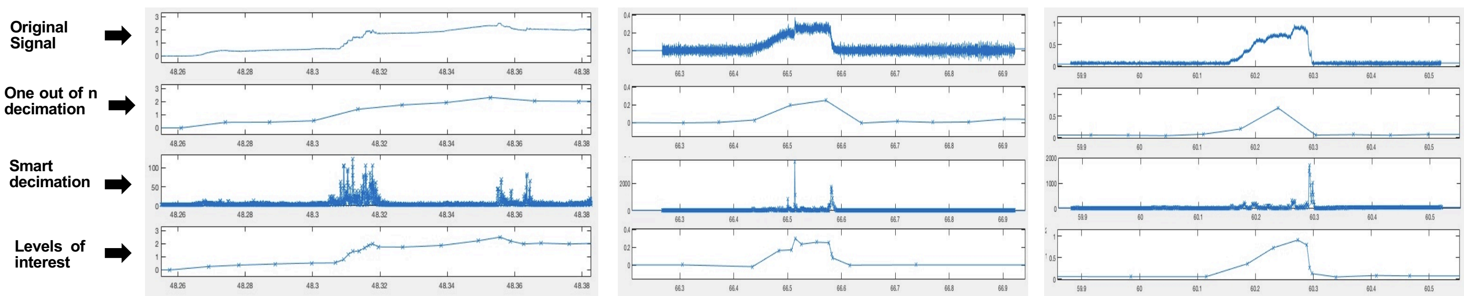
Test results

Cases: Decimation method has been tested with real TJ-II acquired signals. In order to have longer signals, data from consecutive experimental pulses (from 46000 until 46100) have been concatenated.
Evaluation function: To evaluate the quality of a decimated signals respect original signal, two functions have been applied: squared Euclidean distance and correlation coefficient

- **Signal**: DENCM0_ **Pulses**: from 46000 until 46039
- **N Samples**: 13107200 Sampling Rate: 1Mhz
- **Number of decimated points**: 1000

- **Signal**: BOL1 **Pulses**: from 46000 until 46100
- **N Samples**: 6619136 **Sampling rate**: 100KHz
- **Number of decimated points**: 1000

- **Signal**: RX105 **Pulses**: from 46000 until 46100
- **N Samples**: 5049984 **Sampling rate**: 78125 Hz
- **Number of decimated points**: 1000



	Step decimation	Smart decimation
Eucl. Distance	161.07	109.73
Coef. Correlation	0.9969	0.9986

	Step decimation	Smart decimation
Eucl. Distance	107.54	67.23
Coef. Correlation	0.6727	0.8914

	Step decimation	Smart decimation
Eucl. Distance	187.55	99.06
Coef. Correlation	0.7732	0.9421

Acknowledgements

TJ-II operation and experimental teams that have contributed to obtain the experimental data that have been used in this work. Particular thanks are due to Dr. A Baciero, Dr. M Ochando and Dr. F. Medina, for their support and knowledge on tested signals.

References

[1] Plenary oral: J. Vega, "Automatic recognition of plasma relevant events: implications for ITER", 12 IAEA TM on Control, Data Acquisition and Remote Participation for Fusion Research (May 14th 2019)
[2] J. E. Everett, "The Exponentially Weighted Moving Average Applied To The Control And Monitoring Of Varying Sample Sizes", WIT Transactions on Modelling and Simulation, Volume 51, Pages 3 – 13, 2011

Conclusions

- Smart decimation method has been developed as an alternative to the classical "one out of n" decimation mechanism. The new method has been applied to time evolution signals and It is based on level of interest analysis of signal intervals.
- The new smart decimation method has been compared versus the classical step-based decimation method. Both have been tested with real experimental time evolution signals from TJ-II stellarator. The test results show, for all cases, quite better fit to original signal in case of smart decimation method.
- The level of interest of signal intervals is valuable metadata information that can be applied in a wide variety of cases that improves notably the cost effectiveness of experimental data and speeds up data search on big data experiments such as ITER [1].