Fusion Engineering and Design xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Fusion Engineering and Design

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

Automatic feature extraction in large fusion databases by using deep learning approach

Gonzalo Farias[a,*], Sebastián Dormido-Canto[b], Jesús Vega[c], Giuseppe Rattá[c], Héctor Vargas[a], Gabriel Hermosilla[a], Luis Alfaro[a], Agustín Valencia[a]

[a] Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile
[b] Departamento de Informática y Automática, UNED, Madrid, Spain
[c] Asociación EURATOM/CIEMAT Para Fusión, CIEMAT, Madrid, Spain

HIGHLIGHTS

- Feature extraction is a very critical stage in any machine learning algorithm.
- The problem dimensionality can be reduced enormously when selecting suitable attributes.
- Despite the importance of feature extraction, the process is commonly done manually by trial and error.
- Fortunately, recent advances in deep learning approach have proposed an encouraging way to find a good feature representation automatically.
- In this article, deep learning is applied to the TJ-II fusion database to get more robust and accurate classifiers in comparison to previous work.

ARTICLE INFO

Article history:
Received 15 June 2015
Received in revised form 8 June 2016
Accepted 9 June 2016
Available online xxx

Keywords:
Future extraction
Machine learning
Autoencoder
Thomson scattering

ABSTRACT

Feature extraction is one of the most important machine learning issues. Finding suitable attributes of datasets can enormously reduce the dimensionality of the input space, and from a computational point of view can help all of the following steps of pattern recognition problems, such as classification or information retrieval. However, the feature extraction step is usually performed manually. Moreover, depending on the type of data, we can face a wide range of methods to extract features. In this sense, the process to select appropriate techniques normally takes a long time. This work describes the use of recent advances in deep learning approach in order to find a good feature representation automatically. The implementation of a special neural network called sparse autoencoder and its application to two classification problems of the TJ-II fusion database is shown in detail. Results have shown that it is possible to get robust classifiers with a high successful rate, in spite of the fact that the feature space is reduced to less than 0.02% from the original one.

1. Introduction

Experiments in thermonuclear fusion generate a huge number of digital signals. Thousands of signals, sampled at high frequencies, are devoted to studying physical properties of plasma during a discharge producing massive databases.

Nowadays a shot of a tenth of a second could involve 10 GB of data in some fusion devices. However, classifying an input pattern in all storage data could be very hard, and a time-consuming task, without any computer-based assistance. For that reason, a great effort has been done to apply pattern recognition and machine learning techniques to perform pattern classification in fusion databases successfully [1–3]. However, there is still a room for improvements. In particular, feature extraction is still done manually.

Feature extraction is a very critical stage in any machine learning algorithm. Finding attributes of data that are good enough can drastically reduce the dimensionality of the problem and can make any subsequent process of classification and matching easier. On the contrary, the application of poor feature extraction methods could imply a challenge for the following stages, making it very difficult to build a successful pattern recognition system. Despite the importance of feature extraction, the process is commonly done manually by trial and error. Selection of appropriate techniques to obtain useful features from data is not straightforward and normally takes a long time. That is why any advancement in extracting

* Corresponding author.
E-mail address: gonzalo.farias@ucv.cl (G. Farias).

http://dx.doi.org/10.1016/j.fusengdes.2016.06.016

Table 1
TJ-II Thomson Scattering images.

TS Image	Description
BKG	CCD camera background.
STR	Stray light.
ECH	Electron cyclotron resonant heating.
NBI	Neutral beam injection heating.
COF	Cut-off density during electron cyclotron resonant heating.

Table 2
TJ-II waveforms.

Signal	Description
ACTON275	Spectroscopic signal. CV line intensity.
BOL5	Bolometer signal. Total radiation intensity.
DENSIDAD2	Line averaged electron density.
ECE7	Electron cyclotron emission.
GR	First gyrotron.
GR2	Second gyrotron.
HALFAC3	Hα line intensity.
IACCEL1	Neutral beam injector
RX306	Soft X-ray

suitable attributes automatically is very helpful. Recently, deep learning and auto encoders have emerged as one of most encouraging approaches to this end [4].

The paper is structured as follows. Section 2 introduces the datasets used in this work. In particular, images from the Thomson Scattering diagnostic and temporal evolution waveforms of a fusion device are considered. Section 3 describes the deep learning approach and in particular the attributes of autoencoders. Section 4 discusses the main results when the two datasets are classified by using extracted features with autoencoders. Finally, the main conclusions and future works are discussed.

2. TJ-II stellerator

In order to assess the utility of the automatic feature extraction methods for pattern recognition problems in nuclear fusion, we have selected two different classification problems from the TJ-II databases.

The TJ-II is a medium-size nuclear fusion experimental and stellerator device located at CIEMAT in Madrid (Spain). The plasmas in TJ-II are produced and heated with ECRH (2 gyrotrons, 300 kW each, 53.2 GHz, 2nd harmonic, X-mode polarization) and NBI (300 kW). In TJ-II a typical discharge lasts between 150 and 250 milliseconds, and depending on the sampling rate, the number of samples could be in the range of 4000–16000 per shot.

As mentioned before, two types of pattern recognition problems will be discussed in this work. Firstly, the classification of five types of TJ-II Thomson Scattering images, and secondly, the recognition of nine different classes of TJ-II waveforms are taken into account in order to evaluate the performance of the deep learning approach in fusion. Both problems are briefly introduced in the following subsections.

2.1. Thomson scattering images

The Thomson Scattering (TS) diagnostic of the stellarator TJ-II provides temperature and density profiles of plasma. The diagnostic acquires five types of images (spectra of laser light scattered by plasma): CCD camera background (BKG), measurement of stray light without plasma or in a collapsed discharge (STR), images during electron cyclotron resonant heating (ECH), during neutral beam injection (NBI), and after reaching the cut-off density during electron cyclotron resonant heating (COF). Table 1 describes the five classes considered.

An automatic image classification system based on support vector machines (**SVM**) has been in operation for years in the TJ-II Thomson Scattering diagnostic [1,2]. SVM is a very useful method for general-purpose pattern classification. In a few words, given a set of input vectors, which belong to two different classes, SVM maps the inputs into a high-dimensional feature space through some nonlinear mapping (a kernel function), where an optimal separating hyper-plane is constructed in order to minimize the risk of misclassification. The hyper-plane is determined by a subset of points of the two classes, called **support vectors** [5]. SVM can be easily extended for a multi-class problem [2]. We have selected the TS image classification problem to show in detail preliminary results of applying the deep learning approach to image fusion databases. The experiment was carried out with 242 TS images.

Previously to extract features in an automatic way, we apply a pre-processing to all TS images. To this end we have selected Wavelet transform in order to reduce the dimensionality of the problem. According to previous works [1,2], the Wavelet parameters are defined as follows. Wavelet coefficient: approximation; Wavelet mother: Haar; Level of decomposition: 4. As a consequence of the application of Wavelet transform the image dimensions were reduced from the original 576 × 385 (221760) pixels to only 36 × 25

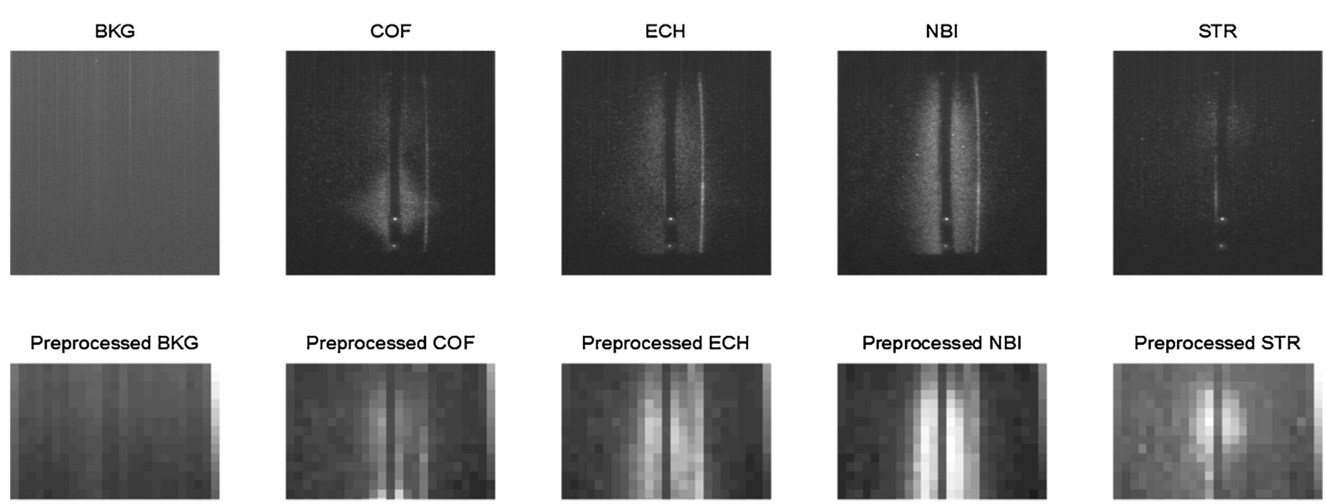

Fig. 1. The upper plots show the five classes of the TJ-II Thomson Scattering images. Lower plots show the corresponding pre-processed versions of the five classes.

Please cite this article in press as: G. Farias, et al., Automatic feature extraction in large fusion databases by using deep learning approach, Fusion Eng. Des. (2016), http://dx.doi.org/10.1016/j.fusengdes.2016.06.016