Computer Physics Communications 185 (2014) 1669–1680

Contents lists available at ScienceDirect

Computer Physics Communications

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

Clustering of periodic multichannel timeseries data with application to plasma fluctuations

S.R. Haskey *, B.D. Blackwell, D.G. Pretty

Plasma Research Laboratory, Research School of Physics and Engineering, The Australian National University, Canberra, ACT 0200, Australia

ARTICLE INFO

Article history:
Received 11 November 2013
Received in revised form
3 March 2014
Accepted 10 March 2014
Available online 17 March 2014

Keywords:
Clustering
Periodic datamining
Von Mises distribution
Magnetic fluctuations
Plasma physics

ABSTRACT

A periodic datamining algorithm has been developed and used to extract distinct plasma fluctuations in multichannel oscillatory timeseries data. The technique uses the Expectation Maximisation algorithm to solve for the maximum likelihood estimates and cluster assignments of a mixture of multivariate independent von Mises distributions (EM-VMM). The performance of the algorithm shows significant benefits when compared to a periodic k-means algorithm and clustering using non-periodic techniques on several artificial datasets and real experimental data. Additionally, a new technique for identifying interesting features in multichannel oscillatory timeseries data is described (STFT-clustering). STFT-clustering identifies the coincidence of spectral features over most channels of a multi-channel array using the averaged short time Fourier transform of the signals. These features are filtered using clustering to remove noise. This method is particularly good at identifying weaker features and complements existing methods of feature extraction. Results from applying the STFT-clustering and EM-VMM algorithm to the extraction and clustering of plasma wave modes in the time series data from a helical magnetic probe array on the H-1NF heliac are presented.

1. Introduction

The identification and characterisation of plasma wave modes as a function of machine and plasma parameters is a subject of considerable interest for plasma magnetic confinement devices. As has been observed with Alfvén waves [1], high energy fusion alphas or neutral beam injection ions can interact with these modes, severely degrading their confinement and driving the modes to large amplitude [2]. This causes significant problems such as damage to the first wall [3], and may prevent fusion plasmas from reaching ignition. Diagnostics such as arrays of magnetic probes are critical for identifying and characterising the spectral and spatial nature of these modes. These diagnostics are "always on" on major experiments, generating extremely large databases of timeseries data which provides a perfect opportunity for knowledge discovery using datamining techniques.

Data clustering, a recognised technique for unsupervised classification, has recently been applied to the field of plasma physics for intelligent data retrieval from large fusion device databases [4–7] and for the identification and classification of wave modes [8–10] using non-periodic clustering algorithms.

The techniques described in this paper address the problem of dealing with periodic data and can be applied to many applications where multichannel diagnostics produces periodic signals. Applications within plasma physics include interferometers, soft X-ray arrays, arrays of magnetic probes and imaging diagnostics. For simplicity we will focus on the application to magnetic probe signals where the spatial information, such as mode numbers, is encoded in the phase differences between magnetic probe signals at the frequency of the mode. These phase differences ($\Delta\psi$) are periodic, $(-\pi, \pi]$, causing problems with standard clustering techniques. Additionally, the number of probes available is often quite large giving rise to high dimensional data. These constraints require the application of specialised clustering techniques. Two options that have good memory scalability are a periodic version of the k-means algorithm and expectation maximisation (EM) using mixtures of multivariate independent von Mises distributions (EM-VMM). Minimal information is available in the literature about the application of EM to multivariate independent von Mises distributions with more than 3 variables, so this is described in detail in Section 4. Previously [8–10], clustering on timeseries data was performed using standard non-periodic clustering techniques by trigonometrically encoding the data ($\sin(\Delta\psi)$ and $\cos(\Delta\psi)$). This method has several drawbacks including artificially creating structure, encoding systematic errors in the data, and doubling the dimensionality of the problem.

* Corresponding author. Tel.: +61 428778853.
E-mail address: shaun.haskey@anu.edu.au (S.R. Haskey).

http://dx.doi.org/10.1016/j.cpc.2014.03.008
0010-4655/

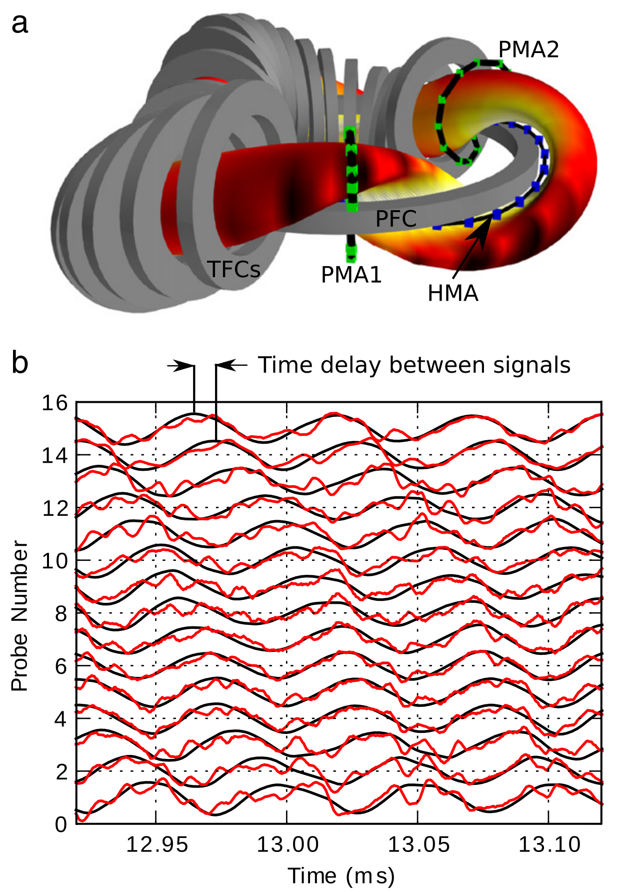

Fig. 1. (a) An overview of the H-1NF heliac including a subset of the equilibrium magnetic field coils (poloidal field coil (PFC), toroidal field coils (TFC)), the poloidal Mirnov arrays (PMA1, PMA2) and the helical Mirnov array (HMA). The surface colour represents the equilibrium magnetic field strength on the last closed flux surface. (b) Examples of the timeseries signals from the probes in the HMA, red is the raw signal and black is a bandpass filtered signal. The time delay in the signal between channels can be converted to a phase difference which represents the spatial structure of the mode. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Using several artificial datasets, we find that the EM-VMM algorithm performs better than the other available algorithms without incurring a significant computational cost. As a case study we successfully apply the EM-VMM algorithm to real data from the H-1NF heliac [11].

H-1NF is a three field-period helical axis stellarator with major radius $R = 1$ m and average minor radius $\langle r \rangle \approx 0.2$ m. The design of the machine allows access to an extensive range of magnetic configurations, making H-1NF well-suited to explore the relationship between plasma behaviour and magnetic configuration [12]. A variety of magnetic fluctuations have been observed with a recently installed helical Mirnov array (HMA) [13], which provides our experimental datasets in this paper. An overview picture of H-1NF including a subset of magnetic field coils and magnetic probe arrays as well as an example of the time trace signals from the HMA when a strong mode is present are shown in Fig. 1(a) and (b) respectively.

Additionally, a pre-processing technique for more robust identification of fluctuations in multichannel oscillatory timeseries data is described. The technique involves a combination of singular value decomposition (SVD) analysis, and an averaged short time Fourier transform followed by clustering (STFT-clustering). The STFT-clustering technique involves finding spectral features using the averaged short time Fourier transform followed by preliminary periodic clustering analysis to identify interesting features.

This paper is organised as follows: Section 2 provides an overview of the feature extraction and clustering process. Section 3 describes the STFT-clustering technique and how combining this with the SVD technique identifies features other techniques miss. Section 4 describes in detail how to apply the expectation maximisation algorithm to a mixture model of multivariate independent von Mises distributions. Section 5 compares the results of applying the periodic and standard clustering techniques to artificial data, and Section 6 shows results from applying STFT-clustering and EM-VMM to experimental data from the H-1NF heliac. Finally we provide some conclusions in Section 7.

2. Overview of the feature extraction and clustering process

For our application, we are ultimately interested in the physical nature of instabilities in plasmas, in particular, their dispersion relations. This information allows us to identify measures that can be taken to prevent these instabilities from growing to destructive amplitudes, and provides information on possible ways to use them beneficially.

Many different types of instabilities give rise to observable fluctuations in a magnetised toroidal plasma, for example ($n = 4$, $m = -3$) global Alfvén eigenmode (GAE), $(5, -4)$ GAE, etc. [1]. Their existence and aspects of their behaviour such as frequency depend on the experimental conditions and plasma parameters such as magnetic field strength and its rotational transform profile and the plasma density. For clustering purposes we assume the spatial structure of a fluctuation instance is what defines it and makes it unique from other fluctuations. Unless the plasma equilibrium is very steady, if a fluctuation exists in a shot it will have different frequencies at different times depending on the plasma parameters such as density and magnetic field strength. Therefore, frequency is not a good identifier of a particular fluctuation and is not used in the clustering process. This and other attributes of each fluctuation instance (time, plasma parameters etc.) are only used later in interpreting the nature of each cluster. Each cluster represents a collection of measurements of the same type of fluctuation that have existed during different experiment conditions which together provide a great deal of information important for interpretation of the underlying physical phenomena.

An overview of the feature extraction, clustering, and analysis process is shown in Fig. 2. The measurements available to identify these instabilities generally consist of timeseries data from arrays of experimental diagnostics such as magnetic pickup probes or multichannel interferometers. In this paper we will focus on magnetic probes but the same technique has been successfully applied to interferometer data.

The magnetic probe signal from a mode that consists primarily of one component such as a global Alfvén eigenmode [1] can be described as follows:

$$V_i \propto \cos(n\phi_{B,i} + m\theta_{B,i} - \omega t). \tag{1}$$

Here, ω is the mode frequency, m represents the poloidal mode number, n the toroidal mode number, i an index in the toroidal array of probes, and, $\phi_{B,i}$ and $\theta_{B,i}$ are the toroidal and poloidal Boozer angles [14,15] of the ith probe, respectively. Examples of the time trace signals from a magnetic probe array due to a mode are shown in Fig. 1(b).

From Eq. (1) we can see that spatial information we are interested in (n and m) is contained in the phase structure of the signal at the frequency of the perturbation (ω). Therefore, the first task is to identify the frequencies of the perturbations over discrete time intervals, and extract the phase structure of the signal at those frequencies for each of the magnetic probes in the array. To make the data independent of the choice of time origin, we calculate the phase difference between successive coils in the array. This forms a