Studies of MHD Stability Using Data Mining Technique in Helical Plasmas

Satoshi YAMAMOTO, David PRETTY[1)], Boyd BLACKWELL[2)], Kazunobu NAGASAKI, Hiroyuki OKADA, Fumimichi SANO, Tohru MIZUUCHI, Shinji KOBAYASHI, Katsumi KONDO[3)], Ruben JIMÉNEZ-GÓMEZ[1)], Enrique ASCASÍBAR[1)], Kazuo TOI[5)] and Satoshi OHDACHI[5)]

Institute of Advanced Energy, Kyoto University, Uji 611-0011, Japan
[1)]Laboratorio Nacional de Fusión, Asociación Euratom-CIEMAT, Madrid 28040, Spain
[2)]Australian National University, Canberra ACT 0200, Australia
[3)]Graduate School of Energy Science, Kyoto University, Uji 611-0011, Japan
[4)]National Institute for Fusion Science, Toki 509-5292, Japan

(Received 5 December 2008 / Accepted 19 November 2009)

Data mining techniques, which automatically extract useful knowledge from large datasets, are applied to multichannel magnetic probe signals of several helical plasmas in order to identify and classify MHD instabilities in helical plasmas. This method is useful to find new MHD instabilities as well as previously identified ones. Moreover, registering the results obtained from data mining in a database allows us to investigate the characteristics of MHD instabilities with parameter studies. We introduce the data mining technique consisted of pre-processing, clustering and visualizations using results from helical plasmas in H-1 and Heliotron J. We were successfully able to classify the MHD instabilities using the criterion of phase differences of each magnetic probe and identify them as energetic-ion-driven MHD instabilities using parameter study in Heliotron J plasmas.

Keywords: MHD stability, data mining, magnetic probe, energetic-ion-driven MHD instability, helical plasma

DOI: 10.1585/pfr.5.034

1. Introduction

Research on toroidal plasmas aiming at realization of magnetically confined thermonuclear fusion enters a new phase as construction begins of the International Thermonuclear Experimental Reactor (ITER), in which self-heating by alpha particle will occur. Knowledge gained from a database based on the experimental results obtained from several tokamaks has influenced the design of the ITER. An alternative design to the tokamak is the helical system which, in principle, can sustain steady state plasmas because its magnetic configuration is mainly produced by external coil currents. The activity to build a database for helical systems has been initiated through a series of Coordinated Working Group Meeting (CWGM) and has produced an International Stellarator/Heliotorn Profile Database (ISHPDB). It is important to create a database in order to obtain a unified knowledge of helical plasmas, as the different helical devices differ in magnetic configuration (e.g. effective helical ripple and dominant Fourier modes) and plasma parameters. Magnetohydrodynamic (MHD) stability, which affects the global plasma confinement and/or particle transport, is studied using large amounts of data obtained at high sampling frequency from multichannel diagnostics such as magnetic probe arrays, electron cyclotron emission (ECE) measurements and soft-X ray diode arrays. Such large databases require sophisticated methods of analysis.

Data mining techniques [1–3] based on statistics, pattern recognition, artificial intelligence and information technology have been used in the areas of distribution and finance for business. Moreover, data mining techniques are also used in the scientific fields of bio-informatics, astronomy and geology. Data mining methods can extract new information because they are able automatic pick out patterns (relationships between data points and parameters) in large amounts of high-dimensional data. We apply a data mining technique to analyze the fluctuation signals within a large database in order to identify MHD instabilities. Moreover, the entry of information about MHD stability classifications into a database enables us to exactly and quickly investigate the characteristics of MHD stability through parameter studies. The data mining technique used here has been shown to be effective for the analysis of MHD stability in H-1 [4] flexible heliac plasmas [5–7] for the first time. We recently applied the data mining technique to the Heliotron J [8] whose magnetic configurations have magnetic well and low magnetic shear in whole plasma. This is the next step study in order to obtain a unified knowledge and its effectiveness for the MHD stability

author's e-mail: syama@iae.kyoto-u.ac.jp

of helical plasmas.

In this article, we discuss the data mining technique and its effectiveness for the MHD stability analysis from the results of H-1 and Heliotron J. Section 2 introduces how to apply the data mining technique to magnetic probe signals related to the MHD instability with a large amount of data. Discussed are the procedures of data mining: pre-processing and filtering of the data, the main clustering, and visualization for the interpretation of results. Section 3 discusses the results of the MHD stability analysis, especially energetic-ion-driven MHD instabilities in Heliotron J plasmas, through the parameter study. A summary is given in Sec. 4.

2. Data Mining Technique for the MHD Stability Analysis

The data mining process employed here consists of three steps, (1) pre-processing, (2) main algorithm for clustering and (3) visualization for the interpretation of results. In order to analyze the MHD stability in helical plasmas we choose, for our initial dataset, multichannel magnetic probe signals such as those from a poloidal array, which provides spatial information about MHD instability. The numbers of magnetic probes belonging same poloidal array are 28 and 14 channels in H-1 and Heliotron J, respectively.

2.1 Pre-processing for the data mining

The extraction of spatial information about MHD instabilities and the filtering out of noise from raw signals are required for the data mining technique to be effective. In order to achieve time resolution we split each shot into short time segments, where the time window is short compared to the plasma discharge time and long compared to the MHD fluctuation time. For example, data are sampled with 1 MHz, then we usually choose 1024 data points, that is, a time segment of about 1 ms. In this article, we focus on the MHD instabilities, of which frequency and/or amplitude usually change in the scale of Alfvén time, but when bursting oscillation with rapid frequency and/or amplitude changes are occurring, the width of time segment should be adjusted accordingly. We take the singular value decomposition (SVD) of all magnetic probes signals for each time segment to separate different MHD instabilities and remove the small singular values, which contain low signal energy and noise. The matrix S, which consists of $N_c \times N_s$ where N_c and N_s are the number of channel and data points within time segment, are represented by SVD.

$$S = UAV^* \tag{1}$$

where the column of U and V correspond to the spatial (topo) and temporal (chrono) singular vector, respectively. V^* means the conjugate transpose of V and the diagonal elements of A are the non-negative singular values. An example of SVD analysis of magnetic fluctuation signals, chrono power spectra and singular values of magnetic probes of Heliotron J is shown in Fig. 1. Here the plasma is produced by the electron cyclotron heating (ECH) and heated and sustained by the neutral beam injection (NBI). We can see two dominant modes having the observed frequency $f_{exp} \sim 40$ kHz in Figs. 1 (b) and (c), and $f_{exp} \sim 105$ kHz in Figs. 1 (d) and (e). Two chronos C1 and C2, and C3 and C4 have very similar frequency and amplitude, respectively. SVD ideally can divide the travelling wave such as rotating mode in the steady state with single helicity into two singular values. Chronos C1 and C2 are the orthogonal sine and cosine components of observed MHD instability, which have very similar frequency and amplitude under the condition that the coherence γ between C1 and C2 is high (e.g. $\gamma > 0.7$) as well as C3 and C4. Group together singular values with $\gamma > 0.7$, then these groups of singular values α_l (e.g. $\alpha_1 = \{a_1, a_2\}$ and $\alpha_2 = \{a_3, a_4\}$) are the data points, where a is the singular value. We take the inverse SVD to get matrix S_l for each fluctuation structure α_l,

$$S_l = UA_lV^*. \tag{2}$$

The rows of the S_l contain the time variation relating the data points for each channel. The power spectra of the topos have a peak at a frequency ω_l. The phase difference $\Delta\psi_{X,Y}$ $(\omega = \omega_l)$ between nearest neighbor each channels X and Y are mapped to the coordinates in $\Delta\psi$-space with $2N_c$ dimensions.

In this study, decomposition of the mode using SVD can be done successfully because the mode is thought to be global Alfvén eigenmode (GAE) having single toroidal and poloidal mode, that is single helicity. If the mode has multi helicity such as toroidicity-induced Alfven eigenmode (TAE), SVD may not be able to decompose the mode into two singular values corresponding to the sine and co-

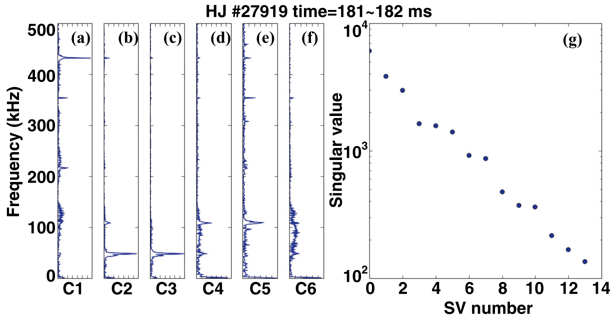

Fig. 1 Typical example of (a)~(f) power spectra of each chrono and (g) singular values in the NBI-heated plasmas of Heliotron J where the line averaged electron density $\langle n_e \rangle \sim 0.7\times10^{19}$ m^{-3} and injected port-through NBI power $P_{inj} \sim 0.5$ MW. C0, C1, ..., C5 are the chronos of the singular value 0, 1, ..., 5 as shown in left figure. There are two distinct modes having the frequency $f \sim 40$ kHz for SV1 (C1) and SV2 (C2), and $f \sim 105$ kHz for SV3 (C3) and SV4 (C4).

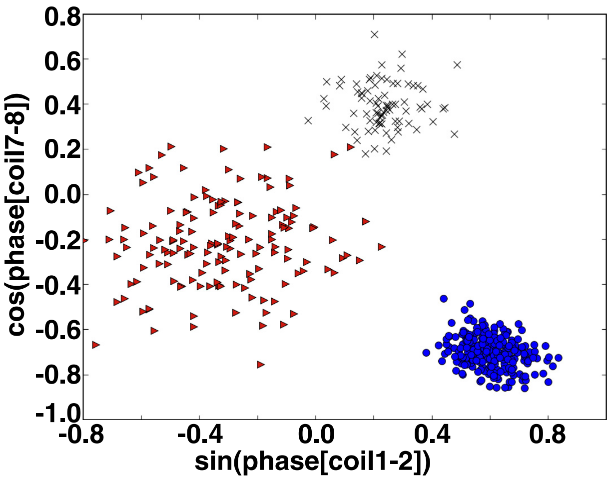

Fig. 2 $\Delta\psi$-space constituted by the phase differences of sine and cosine components for each nearest neighbor magnetic probes pairs of H-1. There are three collectives with different symbol and color. Each collective would pick out by clustering method using the expectation maximization (EM) algorithm.

sine component of the mode. In such case, we have to apply other methods to decompose the mode.

2.2 Clustering

In the $2N_c$ dimensional $\Delta\psi$-space, a class of fluctuation of distinct MHD instability is localized as shown in Fig. 2 from H-1 database. In the H-1 flexible heliac, the plasmas included in our database are produced and heated by the ion cyclotron range of frequency (ICRF) and coherent MHD instabilities driven by the energetic ions were observed [5]. In Fig. 2, we can see three collectives with different symbol and color in a $\Delta\psi$-space consisted of cosine component in channel #7 and #8, and sine component in channel #1 and #2 of magnetic probe. Sine and cosine components of nearest neighbor coil phases are used to overcome the problem of finding a suitable metric in space with 2π periodicity. Therefore, the $\Delta\psi$-space used for H-1 and Heliotron J has 56 and 28-dimensions, respectively. We use the expectation maximization (EM) clustering algorithm, which finds the most likely value of latent variables in a probabilistic model [9]. Here we assume that each type of fluctuation can be described by a $2N_c$-dimensional Gaussian distribution in $\Delta\psi$-space with mean and standard deviation for each cluster as the set of latent variables. The EM method consists of two steps, expectation step and maximization step, and the process is repeated until convergence criterion is reached. The expectation step calculates an expectation of the likelihood by including the latent variables, while maximization step calculates the maximum likelihood estimates of the parameters by maximizing the expected likelihood found on the expectation step.

2.3 Visualization

To visualize and check the result is important because choosing the best number of clusters is very difficult to automate and the automatically calculated results may have errors caused by insufficiency and/or mistake of parameter settings for clustering. In the EM methods employed for this study, we have to optimize the parameter for both numerical scheme (e.g. convergence criterion) and maximization of expectation (e.g. Gaussian width).

Figure 3 shows a *dendrogram*, or *cluster tree* of results from H-1 data, which displays the cluster for each N_{cl} meaning the number of cluster, below some maximum value $N_{cl,\max}$, with all clusters for a given N_{cl} forming a single column. Each cluster plot shows fluctuation frequency vs. κ_h, which is an experimental parameter proportional to the rotational transform. The root of the tree at $N_{cl} = 1$ includes all data points in the scan experiment of rotational transform. Each child cluster is mapped to the cluster on the parent level with which it shares the largest common subset of data points. Cluster branches which do not fork over a significant range of N_{cl} are deemed to be well defined, and the point where well-defined cluster start to break up suggests that N_{cl} it too high.

3. Application Results of Data Mining and Parameter Studies

In this section, we discuss the results of our application of the data mining technique to the Heliotron J plasmas. We analyzed 3786 shots where all of 14 magnetic probes were acquired in the same conditions of isolated amplifier and analog-to-digital convertor. The signals of all magnetic probes in this analysis have been acquired at 1 MHz sampling frequency, and 1024 data points were used for the short time segments ($\Delta t \sim 1$ ms). The size of database with multiple data points per time interval exceeds 2.5 million data points, including the pre- and post-discharge, and dud-discharge (no plasma data). In our database, the magnetic field B_t is set as $B_t = 1.25$ T and plasma is produced by 2nd harmonic resonance of 70 GHz electron cyclotron heating (ECH) in Heliotron J. The plasma is heated by only ECH or only neutral beam injection (NBI) or combination heating of ECH and NBI with discharge duration $t = 120{\sim}150$ ms. In order to retain only significant data points, which correspond to plasma discharge, we considered the H_α signal for each data point produced by the pre-processing discussed above. We applied the filter related to the intensity of H_α signals to the database as shown in Fig. 4 (a) where almost all plasmas are produced by ECH at $t \sim 165$ ms. Shown in Fig. 4 (b) is the removed noises, containing signals caused by non-plasma sources. The data without time dependence (e.g. frequency $f \sim 210$ kHz, 440 kHz, . . . , and so on) in Fig. 4 (a) is coherent electromagnetic noise related to the operation of experiment. Moreover, we have also removed the data below 2 kHz in order to reduce the size of the

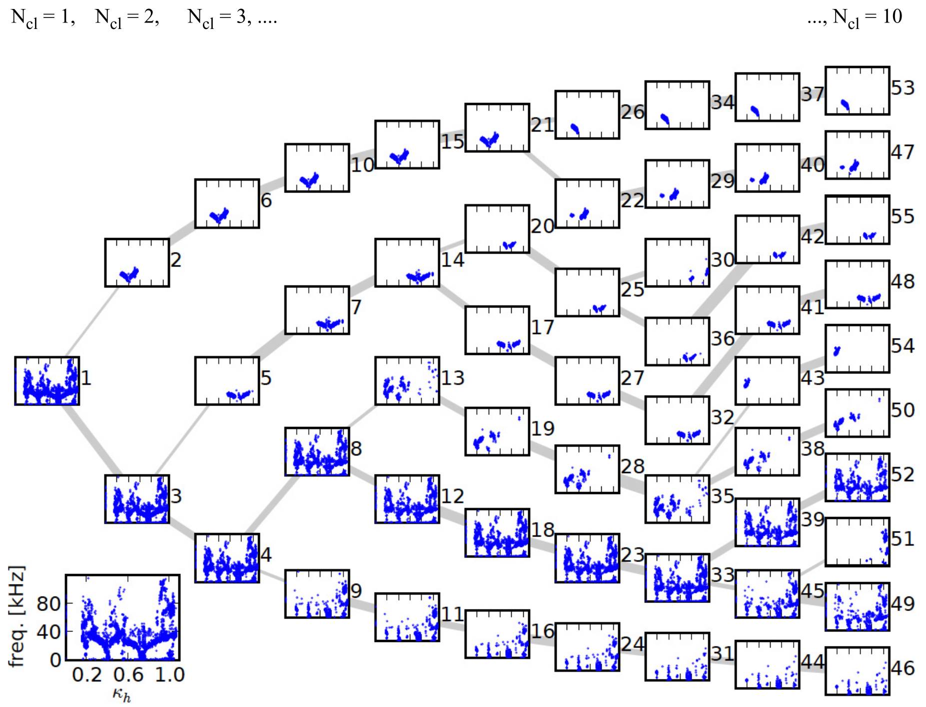

Fig. 3 An Example of cluster tree for MHD instabilities observed in H-1. In all figures, the data are plotted in the rotational transform specified helical coil parameter κ_h and frequency space.

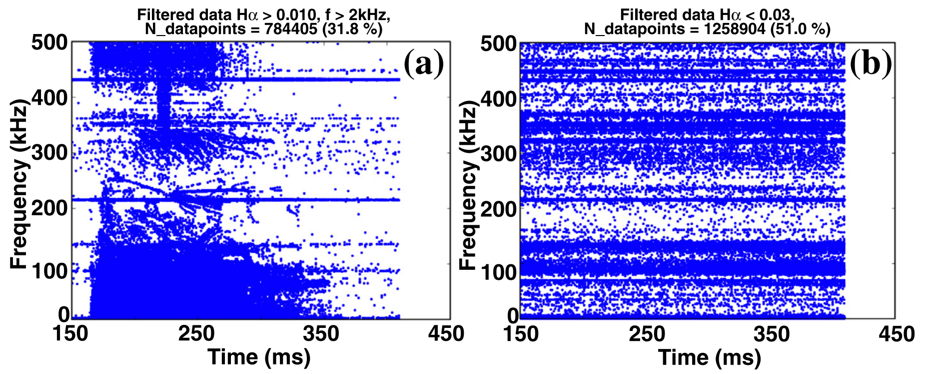

Fig. 4 (a) All data points with high coherence $\gamma > 0.7$ plotted as time (ms) and frequency (kHz) after H_α filtering out process. (b) Removed noise contained the pre-, post- and dud-discharge.

dataset for clustering, retaining the higher frequency fluctuations, some of which have previously been identified as Alfvén eigenmode destabilized by the energetic ions with Alfvénic velocity produced by NBI heating [10]. There are some difficulties to apply the SVD to low frequency fluctuation because their frequency spectra are broadband. Figure 5 shows nine clusters, which are well-defined by the data mining technique using EM algorithm. It seems that the data plots having frequencies $f = 210$ kHz and 430 kHz without time variation in Fig. 5 (f) are not MHD instabilities but electromagnetic noise. The frequency of MHD instabilities in each cluster shown in Fig. 5 is in the range of frequencies $f = 20{\sim}150$ kHz which is close to the frequency of shear Alfvén continua with positive low-m and n (m and n the poloidal and toroidal mode numbers). Figure 6 shows the phase differences between each magnetic probe for each cluster shown in Fig. 5 calculated using the fast Fourier transform (FFT). Each label (a)~(i) in Fig. 6 corresponds to those in Fig. 5. The estimated poloidal mode number m for each cluster is indicated in Fig. 6. Although the direction of propagation of the observed mode is different, the observed modes have low poloidal mode numbers of $|m| < 4$. Here, the sign of poloidal mode number corresponds to the propagation direction of the mode with $m > 0$ and $m < 0$ respectively corresponding to the electron- and ion-diamagnetic drift directions for the condition of $B_t > 0$. To investigate why the observed high frequency mode propagates in

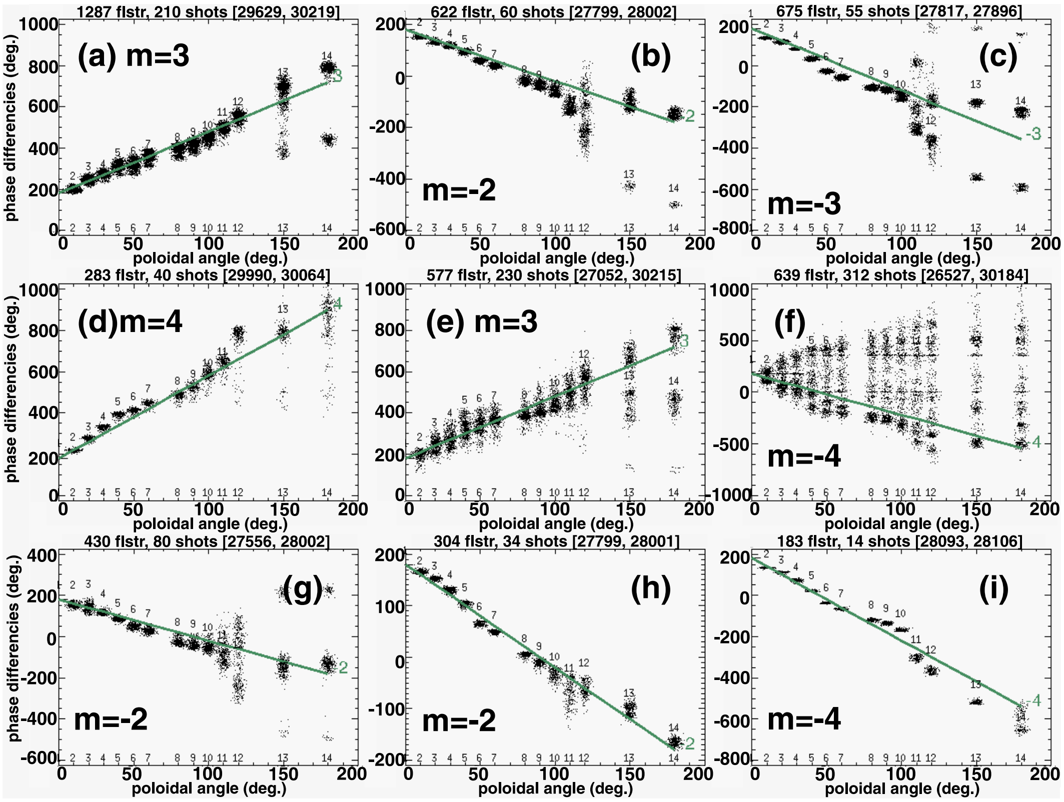

Fig. 6 Phase differences of magnetic probes are illustrated for nine clusters shown in Fig. 5. The vertical and horizontal axes mean the phase differences between magnetic probe #1 corresponding 180 (deg.) and each magnetic probe, and poloidal angle of each magnetic probe in degree. The number near each point indicates the label of magnetic probe. The estimated poloidal mode number m of each MHD instabilities is also indicated in each figure.

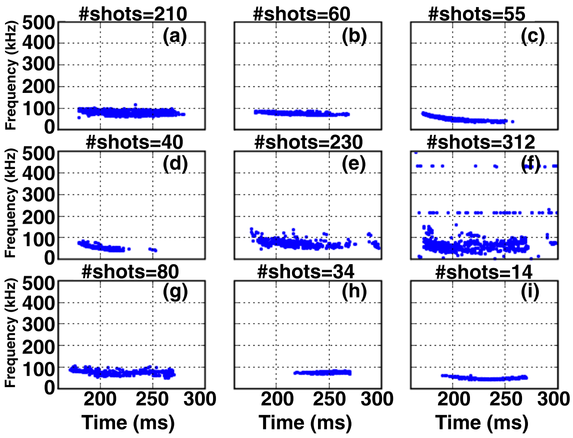

Fig. 5 Nine clusters defined in phase space are shown. Each figure shows the time variation of frequency of data points within each cluster observed in Heliotron J. The number of discharge with observed MHD instabilities belonging each cluster is indicated at the upper of each figure.

the different two directions, we searched the magnetic field strength in the database as show in Fig. 7 (a). As a result of comparison between cluster in Fig. 6 and Fig. 7 (a), almost all of observed modes propagate in the diamagnetic drift direction of ions. The energetic-ion-driven MHD instabilities propagate in the diamagnetic drift direction of ions because the profile of energetic ions, which can resonantly couple with the energetic-ion-driven MHD instabilities, usually has a peak at the plasma center and monotonically decreases toward the plasma edge. The advantage of data mining technique is that it is easy to investigate the characteristics of observed MHD instability through the parameter study. We compared the clusters and the magnetic configuration specified by the edge rotational transform at vacuum and bumpy strength as shown in Figs. 7 (b) and 7 (c), respectively. Although the Heliotron J database does not have a great variety of magnetic configurations, we can see that energetic-ion-driven MHD instabilities are dependent on the configuration. The MHD instabilities in cluster 3 were observed in all magnetic configurations in our database. In previous study in Ref. 10, we identified the observed modes as energetic-ion-driven MHD instabilities with m = 2 and 4, which correspond to the clusters 2, 4, 6, 7, 8, and 9, that is, the cluster 1, 3 and 5 with m = 3 are newly identified as the energetic-ion-driven MHD instabilities. It is noted that the observed modes with same absolute poloidal mode number $|m|$ belonging to different clusters might be the same MHD instability. This is because the clustering only uses the phase differences and

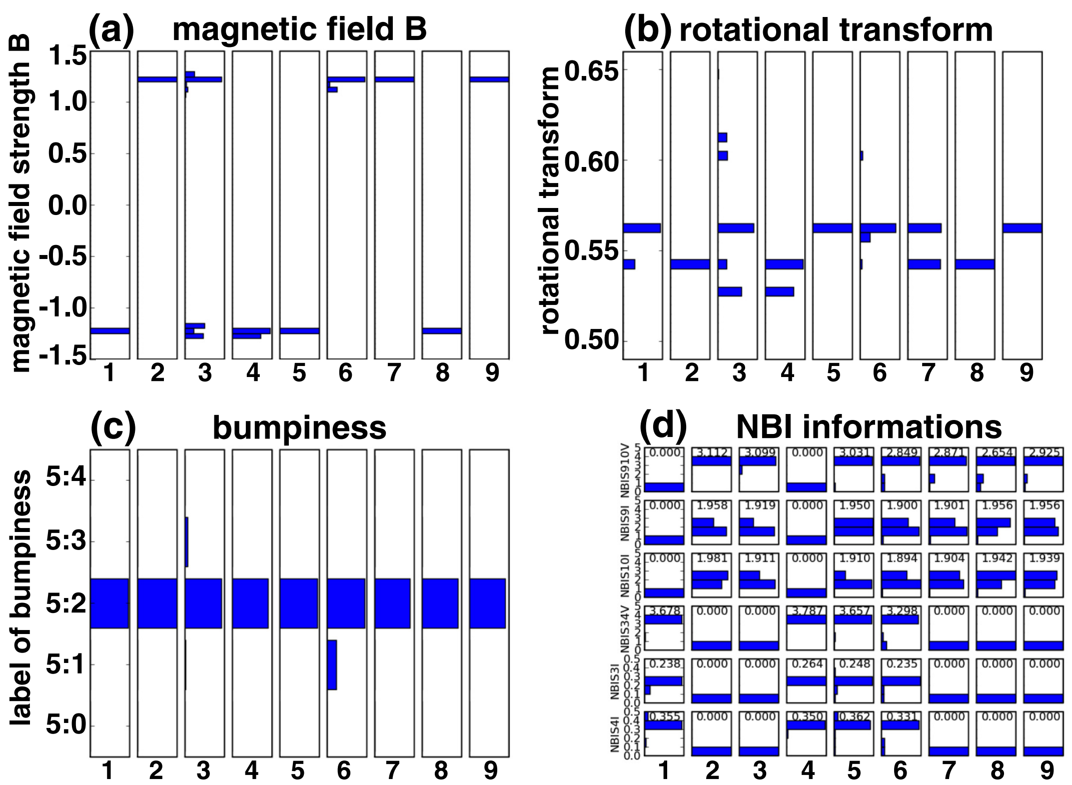

Fig. 7 Parameter dependencies of each cluster of which number showed as horizontal axis corresponds to the number of cluster in Figs. 5 and 6. (a) magnetic field strength B_t for each cluster, (b) rotational transform at edge, (c) label of bumpy field where the value indicates the ratio of coil current for two kinds of toroidal coil, and (d) data of NBIs. NBIS910V and NBIS34V are the acceleration voltage of beam line 1 and 2 of NBIs. NBIS9I and NBIS10I are the beam current of ion sources consisted of beam line 1, respectively. NBIS3I and NBIS4I are the beam current of ion sources consisted of beam line 2, respectively.

therefore will distinguish between opposite poloidal rotation direction determined by the direction of the magnetic field. In order to identify the observed MHD instabilities in each cluster, we investigated the NBI conditions. Almost all observed MHD instabilities were observed in the co-injected NBI heated plasma. From these results we identified the observed MHD instabilities in the nine cluster as energetic-ion-driven MHD instabilities destabilized by the co-flowing energetic ions. In the future plan, we will optimize the pre-processing and clustering algorithm and expand the database regarding to several plasma parameters in order for more clear identification of energetic-ion-driven MHD instabilities and low frequency MHD instabilities which tend to have broadband frequency spectra in Heliotron J.

4. Conclusion

We applied a data mining technique to multichannel magnetic probe signals to analyze the MHD fluctuations in several helical plasmas in order to get unified knowledge of helical plasmas with three-dimensional magnetic configuration. Pre-processing for the data mining technique uses the SVD to search the coherent fluctuations, which would correspond to the MHD instabilities, and to remove noise and unnecessary signals with low intensity. The clustering using EM successfully classified the MHD instabilities from a very large database. We identified the observed MHD instabilities as energetic-ion-driven MHD instabilities such as global Alfvén eigenmodes due to the parameter study using the database including results of the data mining technique and some plasma parameters such as magnetic field strength and rotational transform.

Acknowledgments

The authors would like to acknowledge the H-1, Heliotron J and TJ-II staff. This work was supported by the NIFS/NINS under the NIFS Collaborative Research Program (NIFS04KUHL005, NIFS07KUHL011 and NIFS07KUHL016) and under a project sponsored by the Formation of International Network for Scientific Collaborations.

[1] J. Han and M. Kamber, *Data Mining: Concepts and Techniques* (Morgan Kaufmann, 2001).
[2] W. Frawley and G. Piatetsky-Shairo and C. Matheus, *Knowledge Discovery in Database: An Overview* (AI Magazine, Fall, 1992).

[3] D. Hand, H. Mannila and P. Smyth, *Principles of Data Mining* (MIT Press, Cambridge, MA, 2001).
[4] B.D. Blackwell, Phys. Plasmas **8**, 2238 (2001).
[5] B.D. Blackwell *et al*., Contribution paper of Fusion Energy Conference, EX/P6-3 (2006).
[6] D.G. Pretty and B. Blackwell, Comput. Physics Comm. **180**, 1768 (2009).
[7] D.G. Pretty, PhD thesis, Australian National University (2007).
[8] T. Obiki *et al.*, Nucl. Fusion **40**, 261 (2001).
[9] I.H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques* (Morgan Kaufmann, 2nd edition, 2005).
[10] S. Yamamoto *et al*., Fusion Sci. Technol. **51**, 92 (2007).