Data mining technique for fast retrieval of similar waveforms in Fusion massive databases

J. Vega[1], A. Pereira[1], A. Portas[1], S. Dormido-Canto[2], G. Farias[2], R. Dormido[2], J. Sánchez[2], N. Duro[2], M. Santos[3], E. Sánchez[1], G. Pajares[3]

[1]Asociación EURATOM/CIEMAT para Fusión. Madrid, Spain

[2]Departamento de Informática y Automática. UNED. Madrid. Spain

[3]Departamento de Arquitectura de Computadores y Automática. UCM. Madrid. Spain

Abstract

Fusion measurement systems generate similar waveforms for reproducible behavior. A major difficulty related to data analysis is the identification, in a rapid and automated way, of a set of discharges with comparable behaviour, *i.e.* discharges with "similar" waveforms. Here we introduce a new technique for rapid searching and retrieval of "similar" signals. The approach consists of building a classification system that avoids traversing the whole database looking for similarities. The classification system diminishes the problem dimensionality (by means of waveform feature extraction) and reduces the searching space to just the most probable "similar" waveforms (clustering techniques). In the searching procedure, the input waveform is classified in any of the existing clusters. Then, a similarity measure is computed between the input signal and all cluster elements in order to identify the most similar waveforms. The inner product of normalized vectors is used as the similarity measure as it allows the searching process to be independent of signal gain and polarity. This

development has been applied recently to TJ-II stellarator databases and has been integrated into its remote participation system.

Keywords: Fusion databases; Similar waveforms; Pattern recognition; Data mining; TJ-II

I. Introduction

Fusion devices generate very large databases with a million or more signals and tens or hundreds of thousands of samples per waveform. Moreover, the database contents are not just related to scientific data but also to technical systems. Diagnostics permit the temporal evolution of plasma properties to be followed and "similar" waveforms are generated for reproducible plasma behavior. Control systems record time-dependent signals and similar waveforms characterize analogous discharges.

In general, data analysis in fusion requires searching for "similar" waveforms: statistical analysis, seeking specific behaviours or reviewing previous results. This means selecting a large enough number of signals from different discharges. Such a selection process is usually a manual and tedious procedure in which the signals need to be examined individually.

To automate the searching process for identifying the waveforms that are most "similar" to a reference one, two aspects must be taken into account. The first aspect is the concept of "similar waveforms" itself. Intuitively, one thinks that two signals are similar when one resembles the other. Typically, the identification of similarity in manual searches is carried out by means of visual data analysis. However, several experimental factors connected with signal conditioning (*i.e.* amplification gain and/or signal polarity) may hide the analogous appearance. Thus, the automation of a searching

process implies the definition of a similarity criterion, which requires the introduction of a distance (in the mathematical sense) that can be used to compare how similar two waveforms are. Nevertheless, gains and polarities must be borne in mind as issues.

Once a similarity criterion has been established, the second aspect to be considered is the means to reach for the most similar waveforms. A linear approach might be to compute the similarity factor of a given waveform with all database signals and to sort the waveforms according to the similarity value. However, this procedure is unrealistic in very large databases with a lot of samples per waveform. Therefore, it is necessary to develop methods to reduce the searching space to just the most probable waveforms of being similar.

This article describes a new technique by means of which, given a waveform, it is possible to retrieve similar ones from large databases in a fast and automated way. The method needs an input signal from a researcher and then, it looks for the most similar waveforms within the database. The similarity factor is a measure of how similar the waveforms are in relation to the given one. This factor allows establishing an order in the signals to determine the degree of similarity with the initial one. The technique is based on the development of a classification system that groups waveforms into clusters in accordance with certain rules. Waveform clustering is the essential element to speed up the search and to save computational resources. The searching process is carried out by means of a one by one comparison method but only within those waveforms inside a cluster, rather than by computing the similarity between all database signals.

The technique has been applied to a fusion database. The searching pattern is a full waveform, *i.e.* a waveform that stores the whole plasma evolution during a discharge, from plasma beginning to extinction. A similar waveform recognition system

(SWRS) has been developed for the TJ-II device databases. TJ-II is a medium size stellarator (heliac type) [1] located at CIEMAT in Madrid (Spain). It is a four period device whose main parameters are: $B(0) \leq 1.2$ T, $R(0) = 1.5$ m, $\langle a \rangle \leq 0.22$ m. Two gyrotrons (300 kW each, 53.2 GHz, 2^{nd} harmonic, X-mode polarization) and one NBI (300 kW) provide plasma heating. Presently, the SWRS has been integrated into the TJ-II remote participation system (RPS) [2, 3].

Section II provides an overview of the similar waveform recognition system. Section III describes a general model to develop a very flexible classification system. Finally, section IV explains a specific implementation of a SWRS for the TJ-II environment.

II. Similar waveform recognition system overview

Fusion devices can collect thousands of waveforms per discharge and hence, the database is made up of thousands of signal collections. A signal collection signifies the complete set of recorded signals for an individual waveform for all discharges. For instance, in a Tokamak, the plasma current collection is a collection made up of all plasma current waveforms. The SWRS has been designed to look for similar waveforms within collections. Of course, the present technique would be applicable to all the waveforms of a database but, in a practical environment, the searching process is restricted to waveforms of the same collection.

The waveforms of every signal collection are classified into a series of categories (or clusters). This classification process tries to achieve a convenient set of groups, with a suitable number of waveforms in each cluster, in order to reduce the searching space when looking for similar waveforms for an input signal. The clustering process begins with a feature generation stage to identify measurable quantities that

represent the waveforms with a lower dimensionality. As a result, waveforms are replaced by their feature vectors.

However, creating classifiers involves the use of patterns (from the feature vectors) for learning. Learning refers to some form of algorithm to assign each object to a cluster. There are two common types of learning problems, known as supervised learning and unsupervised learning. Supervised learning is used to estimate an unknown (input, output) mapping from known (input, output) samples. The term 'supervised' denotes the fact that output values for training samples are known. In the unsupervised learning scheme, only input samples are given to a learning system, and there is no notion of the output during the learning [4, 5].

Signal collections can be very different to each other. Thus, several clustering criteria (supervised and unsupervised) could be necessary for optimum classification of the waveforms in each collection. In general, supervised clustering is related to a classification based on physical properties whereas unsupervised clustering is performed when physical criteria are not apparent. Several clustering procedures for fusion experimental signals are discussed in ref. [6].

After building the classification system from feature vectors, the searching process of most similar signals is carried out in four steps. Given a waveform, the first step performs feature extraction. The second one is the classification of the feature vector into one of the existing clusters. The third step is the computation of the similarity factor between the input feature vector and the rest of the cluster feature vectors. Finally, waveforms are sorted according to the similarity measure in descending order.

Recently, other similar waveform recognition systems have been published [7, 8, 9, 10]. The first two are based on Fourier analysis and recognize, on the one hand, slow

varying full waveforms and, on the other hand, patterns within waveforms with, at most, one major frequency component. The searching process of similar waveforms is accomplished by means of an "R-tree" multi-dimensional indexing system. The other two published references are based on the discrete wavelet transform (DWT) [11] and the support vector machine (SVM) learning system. SVM is a universal constructive learning procedure based on statistical learning theory [12]. In references [9, 10], each full waveform is replaced by its wavelet coefficients computed at some decomposition level. To look for similar full waveforms, the wavelet coefficients of a signal are the input to a SVM based learning system. This system identifies the collection that the waveform belongs to and then, a similarity measure is computed between the input signal and all the waveforms (wavelet coefficients) in the collection (it should be noted that no classification system is considered). Two kinds of similarity factors are proposed: Euclidean distance and bounding envelop methods [10].

III. General purpose classification system model

As was mentioned before, any classification process requires a previous feature extraction from the objects to classify*, i.e.* to extract a set of characteristics that represent the object main features. This process is essential for data clustering and it allows the dimensionality of the problem to be reduced considerably.

Because the SWRS can manage several signal collections, a classification system is required for each one. Bearing in mind that each collection may consist of thousands of waveforms and new signals can be incorporated as new discharges are produced, the classification system must follow a very flexible scheme to evolve according to dynamic requirements. To this end, a multi-layer classification system model is proposed. The first layer is made up of the set of clusters that result after the

classification of all waveforms of a collection. Some clusters may contain a high number of waveforms with different patterns and, therefore, they can be sub-classified again. The new clusters form the second layer. The clustering refinement can continue up to reach an optimal classification (fig. 1).

The main properties of the model are:

- A tree structure is generated for each collection. Each cluster is a *node* of the tree. The cluster at the top is the *root* and the clusters at the bottom are the *leaves*.
- The nodes contain waveforms (characterized by their features).
- Each node represents one category (or class) of the classification process performed with the parent node.
- The union of all child nodes is the parent node.
- Different clustering methods can be applied to the several nodes that form a layer.
- The clustering criterion of any node is different from the clustering criterion of any ancestor node.
- Different branches can have different decomposition layers.
- Horizontal expansion: new kind of signals inside a collection can be added as new clusters at any moment in time.
- Vertical expansion: leave clusters can be split in new nodes without affecting the tree structure.

This model enables fine-tuning of the classification system at any moment. Also, it should be noted that the model allows the use of different clustering criteria (supervised and unsupervised) with the several nodes of the tree structure.

IV. SWRS development for the TJ-II database

This section describes the SWRS for the TJ-II environment as well as the TJ-II classification system (feature extraction, clustering method and similarity measure), the client/server architecture for data logging and retrieval, and the integration of the SWRS into the TJ-II remote participation environment.

A. TJ-II SWRS: classification system and similarity measure

The main aim of the classification system is to group the waveforms of a collection into several classes in order to reduce the searching space when looking for similar waveforms. The measurements used for classification are known as features. In the more general case, l features, termed x_i, where $i = 1, 2, \ldots, l$, are used and form the *feature vector*

$$\mathbf{x} = [x_1, x_2, ..., x_l]^T$$

where T denotes transposition. Each feature vector identifies *uniquely* a single object.

Some waveform pre-processing must be performed to build a suitable classification system with the feature vectors. First, it is necessary to bear in mind that searching for similar full waveforms signifies looking for plasmas whose temporal evolution demonstrate similar behaviour. This means that the classification system has to be constructed by referring all signals to the same temporal interval from a single reference (particular event). In medium size devices for example, it is possible to speak about an interval of 400 ms from the plasma start or a 100 ms segment from the beginning of neutral beam injection or a 50 ms interval after an L-H transition. In large devices like JET, time intervals can be several seconds long. In addition, it should be highlighted that different collections can be considered for the same signal. The differences among them are the length of the temporal segments and/or the event that defines the segment start.

The reference time for the TJ-II databases is related to the beginning of the TJ-II discharge. Each discharge has a initial time (τ_0) defined by when the heating starts (ECH or NBI).

In addition, waveform pre-processing is responsible for making the classification system independent of several other factors: signal offset, sampling rates, number of samples or sampling instants.

Waveform pre-processing is made up of 3 stages:

1. **Offset removal**. This stage sets the waveform line base to the 0 V level. This is accomplished by computing the mean value of all samples up to the time instant τ_0, and then subtracting this mean value from all sample in the waveform. This step is essential for the TJ-II first layer clustering criterion as explained below.
2. **Linear interpolation**. In this phase all waveforms are restricted to an interval of 300 ms and signals are aligned with the beginning of the discharges. Linear interpolation of N_I points is carried out between τ_0 – 5 ms and τ_0 + 295 ms, where N_I is the nearest power of 2 (exceeding) to the number of samples in the original waveform in the fore-mentioned interval. The reason for choosing a power of two is related to the application of the wavelet transform for feature extraction.
3. **Feature extraction**. Feature extraction in the TJ-II classification system is achieved by means of the Haar wavelet transform, which has two main advantages. First, it can be computed quickly and easily, and second, the Haar wavelet retains the time and frequency information simultaneously. Feature extraction also allows reducing the problem dimensionality from several tens of thousands of samples to just a few points. Therefore, the waveform obtained in the linear interpolation phase is transformed in accordance with a Haar wavelet

transformation. Different decomposition levels can be chosen (1, 2, 3,…) and a feature vector with a reduced number of characteristics ($N_I/2$, $N_I/4$, $N_I/8$,… respectively) is obtained. Analysis with several decomposition levels were carried out. In conclusion, feature vectors with 256 points allow developing classification systems equivalent to ones built with greater number of features. Waveform pre-processing is summarized in figure 2.

To classify the feature vectors, a supervised clustering criterion is used. The criterion is based on computing the number of features required to reach 99.5% of the feature vector Euclidean norm. In other words, a waveform *w* belongs to cluster K when its feature vector $\boldsymbol{v}_w$ satisfies

$$\sqrt{\sum_{j=1}^{K} v_{w,j}^2} \geq 0.995 \|\mathbf{v}_w\|$$

Therefore, there are 256 possible clusters (as much clusters as features) in the first layer of the classification model. In the present system, no additional sub-classifications have been carried out.

From a physical point of view, the above criterion means that each cluster contains discharges with equivalent pulse lengths. This is a direct consequence of removing the signal offset in the pre-processing stage. After finishing a discharge, the signal level comes back to 0 V and, therefore, the main contribution to the feature vector Euclidean norm takes place during plasma life time.

Taking into account that (1) the waveform processing stage handles a temporal segment of 300 ms and (2) the length of the feature vector is 256, then, the pulse length of two signals in adjacent clusters differs, at most, by 1.172 ms. Therefore, cluster K contains discharges whose length, T_K, satisfies

$$(K-1)\cdot\Delta T < T_K \leq K\cdot\Delta T$$

where ΔT is the maximum difference between signals of adjacent clusters.

As mentioned previously, the searching process of similar waveforms is accomplished among waveforms of a single cluster. However, the T_K parameter can be considered small to completely discriminate similar waveforms in adjacent clusters. Hence, for practical purposes, the searching process is not limited to a single cluster, rather to an odd number (N_C) of them. The search is symmetrically distributed around the initial cluster, covering both sides (left and right) with $(N_C - 1)/2$ clusters each.

A similarity measure must be introduced in order to identify signals that are most similar to a given waveform. When the angle between two vectors is a meaningful measure of their similarity, then the normalized inner product may be an appropriate similarity function. The absolute value of this quantity has been chosen as the similarity measure and the vectors are feature vectors ($\boldsymbol{u}_w$ and $\boldsymbol{v}_w$).

$$S_{\mathbf{uv}} = \left|\cos\alpha\right| = \frac{\left|\mathbf{u}_w \cdot \mathbf{v}_w\right|}{\left\|\mathbf{u}_w\right\| \cdot \left\|\mathbf{v}_w\right\|}, \quad 0 \le S_{\mathbf{uv}} \le 1$$

The absolute value of the normalized inner product provides two main advantages. The method does not depend on either amplification gains (*i.e.* waveforms whose difference is a gain factor are recognized as equal signals) or signal polarity (*i.e.* inverted waveforms are perceived as equal signals).

It should be noted that the present technique retrieves the most similar waveforms to a given one, but it does not imply that the signals are almost equal (similarity near 1). The method finds the most similar signals but the similarity factor can be low (close to 0). In other words, the technique can get a list of signals from the database although the waveforms do not resemble between them. When this happens, the initial signal can be considered as an outlier.

The classification system model and the inner product similarity measure constitute a very powerful recognition system when searching for any kind of waveform. The waveforms, first, are not restricted by signal characteristics (for instance, frequency components) and, second, they do not need to be defined in advance (any kind of signal can be considered a pattern). These facts ensure capabilities to seek for any kind of waveform at any moment. Computational resources are maintained at a minimum. The major requirement is disk storage, although in reality this is not very large.

A first SWRS was developed in a Windows XP Pentium IV computer with the Matlab software package [13]. Single layer classification systems for several collections were created with 846 feature vectors of dimension 256. To test the searching process, one waveform was chosen in a random way and the ten most similar waveforms were retrieved by looking for similarity in 9 clusters ($N_C = 9$). The reason of choosing this value is to compute the similarity between signals whose discharge lengths differ at most in 10 ms (about 3% of the temporal segment, which is 300 ms). It should be taken into account that the temporal difference between adjacent clusters is 1.172 ms.

Of course, the method always finds the initial waveform with a similarity factor of 1. Figure 3 shows results for a collection of bolometry signals. In this case, the 99.5% point of the feature vector norm was reached with 136 coefficients ($K = 136$). The number of waveforms in the 9 clusters was 75 and, therefore, the process computed 75 normalised inner products and sorted them. The total time for computations was 2.283 s and the CPU time was 1.482 s. The figure also gives information on measures of similarity, feature vectors and real waveforms. It should be noted that the similarity criterion is independent of signal polarity. Figure 4 shows results for a soft x-ray signal collection. The CPU time for this calculation was 5.718 s. This example illustrates the

independency of amplification gain and also the capability of finding oscillating patterns.

Note that computation times for the searching process are short enough and there is no need of defining additional layers to the classification system.

B. Client/server architecture for the TJ-II SWRS

An optimum use of the SWRS implies a general development for use in a shared environment (both local and wide area networks). Client/server architecture ensures a suitable means of interaction. The TJ-II SWRS server part resides in the TJ-II central data server (an AlphaServer computer with Tru64 UNIX operating system). Communication protocol between clients and server is TCP/IP, using Berkeley Sockets API (Application Program Interface). The communication mechanism is connection oriented and the server was developed as a concurrent server. This scheme is sufficiently general to service multiple concurrent connections and to allow the development of clients for several platforms and applications: visual data analysis applications and software library development.

The server part is in charge of managing the different classification systems for the several waveform collections (fig. 5). Each classification system is characterized by a database (that stores the feature vectors of the waveforms) and a set of files (one per cluster) containing the shot numbers that belong to the cluster. The database is based on a traditional method to handle queries on primary (*i.e.* unique) keys: hashing. In particular, the ndbm package of the UNIX operating system is used. Feature vectors are indexed according to shot number. The database and the 256 files (C_K, K = 1,…, 256) share a common directory in the AlphaServer computer.

The server part provides computational resources

- to create new classification systems

- to include new data (waveform classification)
- to retrieve information (similar waveform searches)

New data integration is carried out in two steps (fig. 6): (1) data pre-processing and (2) feature vector classification. The former writes the wavelet coefficients into the database and the latter appends the shot number to the corresponding cluster file.

The searching process of similar waveforms takes 4 stages (fig. 6): (1) input signal pre-processing, (2) feature vector classification into a cluster, (3) similarity factor computation with the feature vectors of N_C adjacent clusters and (4) similarity factor sorting in descending order. Note that the searching process does not require that the input signal should be previously classified. The existence of both the cluster files with the cluster composition and the database greatly speeds up the similarity factor computation.

In its first stage, the TJ-II SWRS system was built with four signal collections. The first one corresponds to a central chord of a soft x-ray detector array. The second one represents an integrated radiation signal from a line-of-sight near plasma centre measured by a bolometer detector. The third collection is an Hα emission measurement. Finally, the last collection groups a line averaged electron density measurement. The system has been created with data corresponding to the 2004/2005 TJ-II experimental campaigns. Each collection has 1350 waveforms approximately. Like in the Matlab case, the number of adjacent clusters to look for similar waveforms is 9.

C. Integration into the TJ-II remote participation system

A first application of the SWRS was its integration into the TJ-II remote participation system. A Java application can be downloaded from the corresponding web page, according to the usual procedure in the TJ-II RPS [3]. Security for download and execution is based on the PAPI authentication and authorization system [14]. This

Java application provides researchers with a point and click graphical user interface (GUI) to select waveforms and to search for the most similar ones. The application retrieves at most N_S similar waveforms and it shows the respective similarity factors. At present, N_S is normally 20, but this number can be easily modified at any moment. The GUI incorporates controls for horizontal and vertical expansion of traces, signal variable offset, absolute and relative measurements on waveforms, zoom capabilities and signal display with a variable sampling rates.

Figure 7 shows the GUI after searching for similar waveforms. Signal selection is carried out under the 'Master signal' tab, whereas similar signal display is performed in the 'Similar signals' tab. Two list-boxes in the bottom of the window show similarity factors and shot numbers. The box in the bottom centre provides information about the signals in display. The waveforms inside the box at the bottom right are not displayed. However, signals can be moved between boxes either to appear or to disappear in the graphical area.

Figure 8 shows the communication diagram with both remote and local users. Data exchange protocol is very simple and is carried out by means of Java Server Pages (JSP). The client application can send two types of queries. First, the user asks for a signal name and a shot number to perform visual data analysis for signal selection ('Master signal' tab). The server side transmits the data. Once the waveform to look for similar signals has been established, the user asks for them and the application resends signal name and discharge. Now, the server transfers, on the one hand, the similarity and shot number for N_S waveforms and, on the other hand, the waveforms. Signal transmission is accomplished in compressed format in order to save bandwidth and to speed up the transfer process. Data compression is realized according to standard TJ-II methods based on lossless techniques [15].

V. Acknowledgements

The authors wish to thank Prof. Sebastián Dormido Bencomo (UNED) and Prof. Jesús Manuel de la Cruz (UCM) for their constructive comments and invaluable guidance.

References

[1] C. Alejaldre et al. Plasma Phys. Controlled Fusion 41, 1 (1999), pag. A539.

[2] J. Vega et al. Fusion Engineering and Design, 74 (2005) 775-780.

[3] J. Vega et al. Overview of the TJ-II remote participation system. Fusion Engineering and Design. (in press).

[4] R. O. Duda, P. E. Hart, D. G. Stork. *Pattern classification*, (2nd edition). John Wiley & Sons, INC. 2001.

[5] V. Cherkassky, F. Mulier. *Learning from data*. John Wiley & Sons, INC. 1998.

[6] N. Duro et al. Automated clustering procedure for TJ-II experimental signals. Fusion Engineering and Design. (in press).

[7] H. Nakanishi et al. Fusion Engineering and Design, 71 (2004) 189.

[8] H. Nakanishi et al. Similar Pattern Search for Time-Sectional Oscillation in Huge PlasmaWaveform Database. Fusion Engineering and Design. (in press).

[9] S. Dormido-Canto et al. Review of Scientific Instruments 75, 10 (2004) 4254-4257.

[10] G. Farias et al. Searching for patterns in TJ-II time evolution signals. Fusion Engineering and Design. (in press).

[11] S. Mallat. *A Wavelet Tour of Signal Processing*, (2nd edn), Academic Press, 2001

[12] V. Vapnik. *The Nature of Statistical Learning Theory* (2nd edn.), Springer. 2000.

[13] http://www.mathworks.com.

[14] R. Castro et al. An authorization and authentication infrastructure: the PAPI system. Fusion Engineering and Design. In press.

[15] J. Vega et al. Review of Scientific Instruments 67, 12 (1996) 4154-4160.

Figure captions:

Fig. 1: Classification system model.

Fig. 2: Waveform pre-processing. (a) Initial waveforms. (b) Offset removal and linear interpolation. (c) Signal alignment for feature extraction. (d) Feature extraction: Haar wavelet coefficients.

Fig. 3: SWRS results for a bolometry signal.

Fig. 4: SWRS results for a soft x-ray signal.

Fig. 5: Directory structure in the central server.

Fig. 6: Steps for waveform classification and searching processes.

Fig. 7: TJ-II remote participation system GUI.

Fig. 8: TJ-II remote participation system data flow.

Figure 1

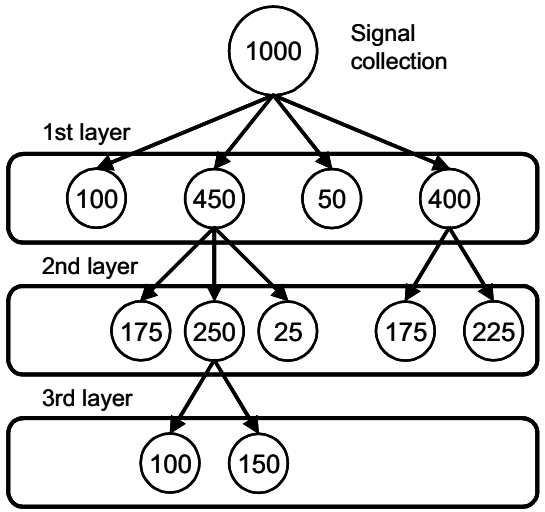

1000
Signal collection
1st layer
100
450
50
400
2nd layer
175
250
25
175
225
3rd layer
100
150

Figure 2

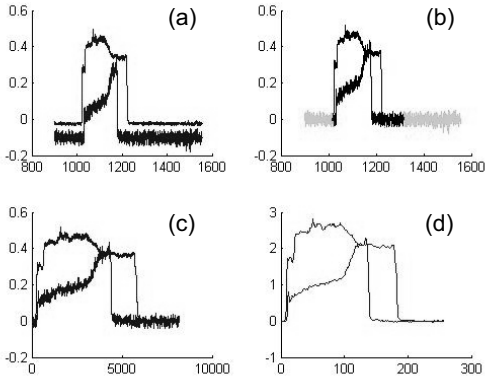

(a)
(b)
(c)
(d)
0.6
0.4
0.2
0
-0.2
800
1000
1200
1400
1600
0
5000
10000
3
2
1
-1
100
200
300

Figure 3

collection	K	feature vectors	time (s)
BOL5	136	75	2.283

Similarity	Shot
1.000000	13774
0.999249	13775
0.997523	13770
0.997363	13764
0.996873	12881
0.996541	13773
0.996502	13760
0.995959	13777
0.995498	13761
0.995408	12957

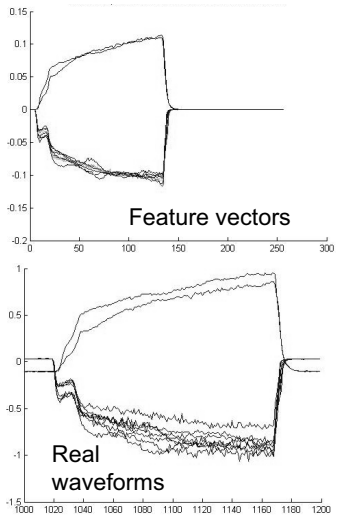


Figure 4

collection	K	feature vectors	time (s)
RX306	178	280	9.394

Similarity	Shot
1.000000	13438
0.997681	13439
0.991461	13435
0.990737	13437
0.988594	13487
0.986054	13355
0.986046	13239
0.985052	12749
0.984527	13202
0.984195	13485

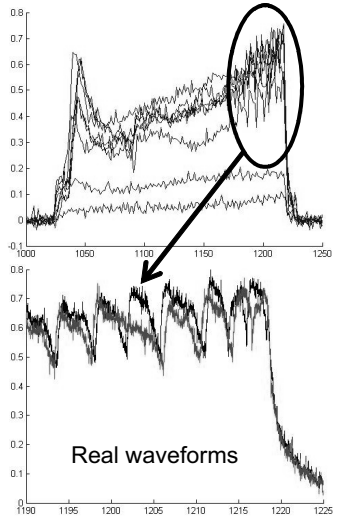


Figure 5

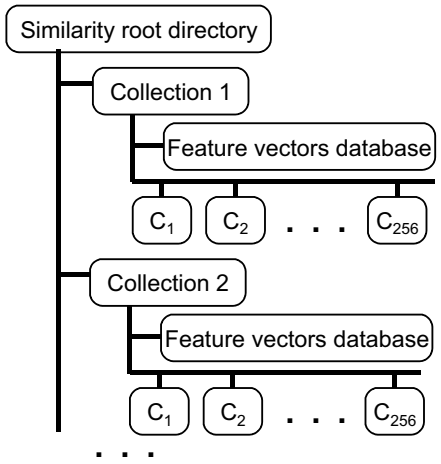

Similarity root directory
Collection 1
Feature vectors database
C1
C2
. . .
C256
Collection 2
Feature vectors database
C1
C2
. . .
C256
. . .

Figure 6

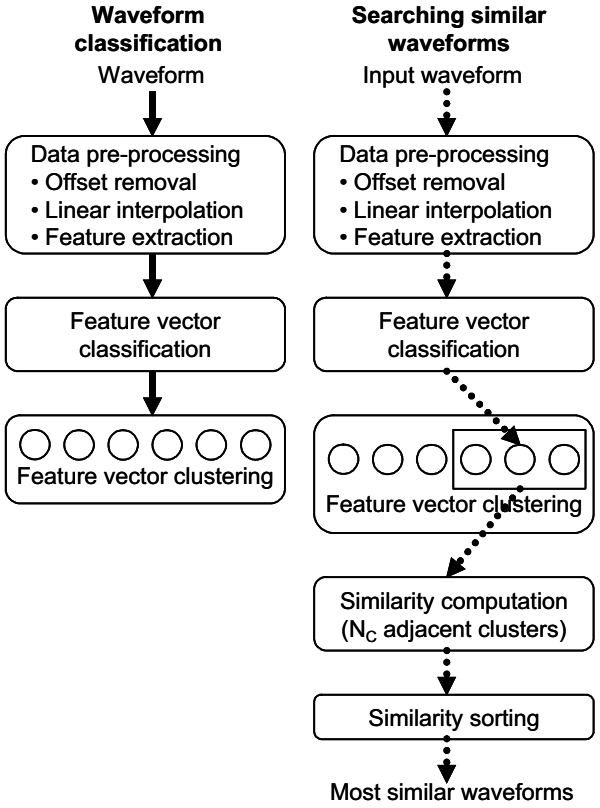

Waveform classification
Waveform
Data pre-processing
• Offset removal
• Linear interpolation
• Feature extraction
Feature vector classification
Feature vector clustering
Searching similar waveforms
Input waveform
Data pre-processing
• Offset removal
• Linear interpolation
• Feature extraction
Feature vector classification
Feature vector clustering
Similarity computation
(N_C adjacent clusters)
Similarity sorting
Most similar waveforms

Figure 7

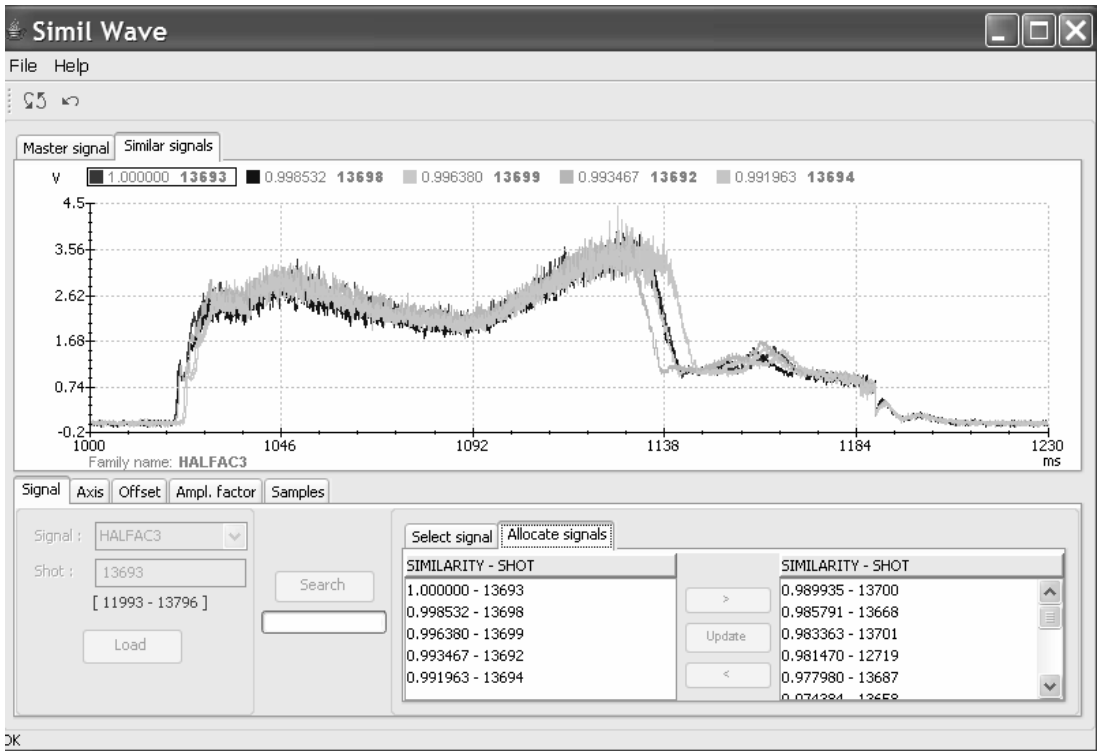
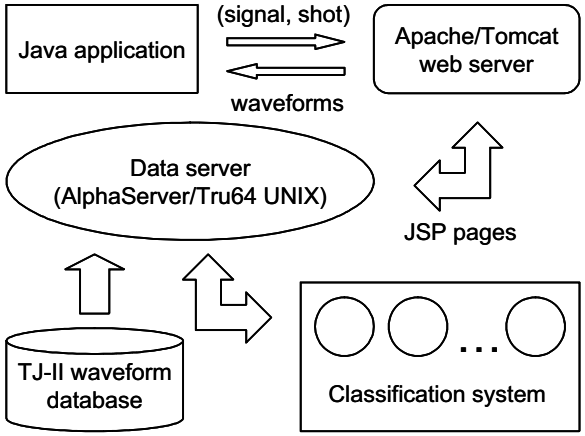

Simil Wave
File Help
Master signal
Similar signals
V
1.000000 13693
0.998532 13698
0.996380 13699
0.993467 13692
0.991963 13694
4.5
3.56
2.62
1.68
0.74
-0.2
1000
1046
1092
1138
1184
1230
ms
Family name: HALFAC3
Signal
Axis
Offset
Ampl. factor
Samples
Signal : HALFAC3
Shot : 13693
[11993 - 13796]
Load
Search
Select signal
Allocate signals
SIMILARITY - SHOT
1.000000 - 13693
0.998532 - 13698
0.996380 - 13699
0.993467 - 13692
0.991963 - 13694
>
Update
<
SIMILARITY - SHOT
0.989935 - 13700
0.985791 - 13668
0.983363 - 13701
0.981470 - 12719
0.977980 - 13687
OK

Figure 8


Java application
(signal, shot)
waveforms
Apache/Tomcat
web server
Data server
(AlphaServer/Tru64 UNIX)
JSP pages
TJ-II waveform
database
Classification system