Fuzzy Decision Tree and Particle Swarm Optimization for Mining of Time Series Data

Maya Nayak Professor Orissa Engineering College, BPUT, Bhubaneswar, India Satyabrata Dash Asst .Professor Orissa Engineering College BPUT, Bhubaneswar, India

ABSTRACT

This paper presents a new approach for power signal time series data mining using S-transform based K-means clustering technique and fuzzy decision tree. Initially the power signal time series disturbance data are pre-processed through an advanced signal processing tool such as S-transform and various statistical features are extracted, which are used as inputs to the K-means algorithm for disturbance event detection. Particle Swarm Optimization (PSO) technique is used to optimize cluster centers which can be inputs to a fuzzy decision tree for pattern classification of time varying database like the power signal data bases.

Index Terms—Time frequency transform, S-transform, Power signal time series data, K-means clustering, decision tree.

1.INTRODUCTION

Cluster analysis has become an important technique in exploratory data analysis, pattern recognition, machine learning, neural computing, and other engineering system studies [1-5]. The clustering aims at identifying and extracting significant groups in underlying data. Each group, called a cluster, consists of objects that are similar to one another and dissimilar to objects of other groups. Data are termed as static if all their feature values do not change with time or change negligibly. The bulk of clustering analysis has been applied to static data. Unlike static data, the features of the time series comprise values that change with time. Time series clustering [6-8] has been very effective in providing pertinent and useful information in various application domains as a part of temporal data mining research. Time series data are of interest because of their pervasiveness in various areas ranging from science, engineering, business, finance, economics, healthcare, etc.

Various algorithms have been developed to cluster different types of time series data. In order to apply the clustering algorithms, it is necessary to convert the time series data to the static one and modify the similarity or distance measures for clustering. The various power line disturbances include voltage sag, swell, impulsive, and oscillatory transients, multiple notches, momentary interruption, harmonics, and voltage flickers, etc. Unfortunately most of the recorders used for power signal data mining rely on visual inspection of data record creating an unprecedented volume of data to be inspected by engineers. Thus an automatic recognition or classification of the voluminous data is required for improving the quality of electricity supply.

Although wavelet multiresolution analysis combined with a large number of neural networks provides efficient classification of power quality (PQ) events [9-12] the time domain featured disturbances such as sags, swells, etc. may not easily be classified. In addition if an important disturbance frequency component is not precisely extracted by wavelet transform which consists of octave band pass filters the classification accuracy may also be limited. This paper, therefore, presents a generalized discrete S-transform [13,14] with modified Gaussian window that varies inversely with frequency for the detection, localization of power line disturbance signal time series database. The modified S-transform simply designated as MST produces a time-frequency representation of a time varying signal by uniquely combining the frequency dependent resolution while simultaneously localizing the real and imaginary spectra. The MST of a Power line disturbance signal time series data provides contours which closely resemble the disturbance patterns unlike the wavelet transform and hence the features extracted from it are very suitable for developing highly efficient and accurate classification scheme. Further the MST analysis of time varying signal data yields all of the quantifiable parameters for localization, detection and quantification of signals comprising a time series database.

After the features are extracted from the time-series data using a modified discrete S-transform, a clustering analysis is used to group the data into clusters and thereby identifying the class of the data. The well known Fuzzy C-means Algorithm [15] is commonly used for data clustering but suffers from trial and error choice of the initial cluster centers and also the noise present in the original time series data. In addition to this the traditional Fuzzy C-means Algorithm gets stuck in the local minima and, therefore, does not offer robustness in clustering the features. In order to overcome this problem a simpler hybrid K-means clustering algorithm along with Particle swarm optimization (PSO) technique [16,17] is used for clustering the features into distinct groups that may be used further for classification. Unlike other evolutionary learning algorithms, PSO needs smaller parameters to decide and can be easily implemented, and has stable convergence characteristic with good computational efficiency.

2. MODIFIED DISCRETE S-TRANSFORM FOR FEATURE EXTRACTION

The S-transform (ST) of a time series x(t) is defined as [1]:

$$S(t,f) = \int_{-\infty}^{\infty} x(\tau) \frac{1}{\sigma(f)\sqrt{2\pi}} e^{\frac{-(t-\tau)^2}{2\sigma(f)^2}} e^{-2i\pi f\tau} d\tau \qquad (1)$$

Thus, an alternative representation for the generalized discrete Stransform with modified Gaussian window is

$$S[j,n] = \sum_{m=0}^{N-1} X[m+n] e^{(-2\pi^2 m^2 \beta^2 / (a+bf^p)^2)} e^{i\frac{2\pi mj}{N}}$$
(2)

Where N is the number of frequency points , and X[m + n] is obtained by shifting the discrete Fourier Transform (DFT) of x(k) by n. The output of the S-transform is an n×m matrix, whose rows pertain to frequency and columns indicate time. Each column thus represents the "local spectrum" for that point in time. From the ST matrix we obtain the frequency-time contours having the same amplitude spectrum and these contours can be used to visually classify the nature of the signal database and its change of frequency and the harmonic content if any during the distortion. Fig.1 shows a block diagram approach for time series data clustering. formats for your particular conference.

3. K-MEANS CLUSTERING

In the field of clustering, K-means algorithm is the most popularly used algorithm to find a partition that minimizes mean square error (MSE) measure. For space S which has the K groups and the N points $\{x_1, x_2, ..., x_M\}$, the definition of the clustering analysis is as follows: In many clustering techniques, the K-means (generally called hard C-means) algorithm is one of well-known hard clustering techniques. It can allocate the data point x_i to the closest cluster center z_i by using Euclidean distances:

$$d = \|x_i - z_j\|, i \in \{1, 2, ..., N\}, j \in \{1, 2, ..., K\}$$
(3)

A good method will cluster data set $X = \{x_1, x_2, ..., x_N\}$ into K well partitions with $2 \leq K \leq M-1$. When we have an unlabelled data set, it is very important to define an objective function for a clustering analysis method. Intuitively, each cluster shall be as compact as possible. Thus, the objective function of the K-means algorithm is created with the Euclidean norm given by:

$$J_{E} = \sum_{j=1}^{K} \sum_{x_{i} \in C_{i}} \left\| x_{i} - z_{j} \right\|^{2}$$
(4)

where z_j is the jth cluster center. The necessary condition of the minimum J is

$$z_{j} = \frac{1}{N_{j}} \sum_{x_{i} \in C_{j}} x_{i} , \quad j = 1, 2, \dots, K$$
(5)

Where N_j is the number of points belonging to cluster C_j .

The K-means clustering process can be stopped when any one of the following criteria are satisfied: when the maximum number of iterations has been exceeded, when there is little change in the centroid vectors over a number of iterations, or when there are no cluster membership changes.

3.1. Particle swarm Optimization

PSO is a novel stochastic origin in the motion of a flock of birds

searching for food. The basic PSO algorithm is started by scattering a number of particles called swarms in the function search apace. Each particle moves in the search space looking for the global minimum or maximum. During its flight each particle adjusts its trajectory by dynamically altering its velocity according to its own flying experience and the flying experience of the other particles in the search space. The PSO approach is becoming very popular due to its simplicity of implementation and its ability to quickly converge to a reasonably good solution. Each particle represents a position in *i* in the *j* dimensional search space, and is flown through this space adjusting its position towards both the particle's own best position found thus far, and the best position in the neighborhood of that particle. For a particle moving in a multidimensional search space let

 $X_{i,j}$ and $V_{i,j}$ denote the position of ith particle in then jth dimension and velocity at time t. The local best position of the ith particle is obtained as

$$pbest_{i}(t+1) = \begin{cases} pbest_{i}(t), \ if \ F(x_{i}(t+1) > F(pbest_{i}(t)) \\ x_{i}(t+1), if \ F(x_{i}(t+1) < F(pbest_{i}(t)) \\ \end{cases}$$
(6)

and the global best position is obtained as

$$\hat{g}best(t) = \min \begin{cases} F(pbest_0(t), F(pbest_1(t), F(pbest_1(t), F(pbest_2(t), \dots, F(pbest_s(t)))) \end{cases}$$
(7)

where F denotes the fitness of the particles and is obtained from the Euclidian distance and in the section 4.2.

The modified velocity and position of each particle at time (t+1) can be calculated as

$$v_{i}(t+1) = K \begin{bmatrix} w(t).v_{i}(t) + c_{1}.rand_{1}.(pbest_{i} - x_{i}(t)_{i}(t)) + \\ c_{3}.rand_{3}.(pos_{r} - x_{i}(t)) \end{bmatrix}$$
(8)

where v_i is the velocity of ith particle at time t+1, x_i is the current position, w is the inertia weight factor, where i = 1, 2, ..., N; c_1 , and c_2 are, respectively, the cognitive, and social factors; rand₁, and rand₂ are random numbers uniformly distributed within [0,1], K is the constriction factor which is a function φ_1 and φ_2 given by

$$K = \frac{1}{\left|2 - \varphi - \sqrt{\varphi^2 - 4\varphi}\right|} \text{, and } \varphi = \varphi_1 + \varphi_2; \ \varphi > 4$$
(9)

3.2. Pso Based Clustering

In the context of clustering, a single particle represents the N_c cluster centroid vectors. That is, each particle x_i is constructed as follows:

$$x_i = m_{i1}, m_{ij}, \dots, m_{iN_c}$$
(10)

where $m_{i,j}$ refers to the j-th cluster centroid vector of the i-th particle in cluster C_{ij} . Therefore, a swarm represents a number of candidates clustering for the current data vectors. The fitness of the particle i is easily measured as the quantization error as

$$F_{i} = \frac{\sum_{j=1}^{N_{c}} \left| \sum_{\forall Z_{p} \in Cij} d(z_{p}, m_{j}) / |C_{ij}| \right|}{N_{c}}$$
(11)

where d is defined in equation (15), and $|C_{ij}|$ is the number of data vectors belonging to cluster C_{ij} .

The K-means algorithm exhibits a faster, but premature convergence due to a large quantization error, while the PSO algorithm gives lower quantization errors and a robust convergence.

$$x_i(t+1) = x_i(t) + v_i(t+1)$$
(12)

Further the position and the velocity of each particle is limited by the following expressions:

$$x_i^{\min} \le x \le x_i^{\max}$$
, and $-V_i^{\max} \le v \le V_i^{\max}$ (13)

Also for the jth dimension, the maximum velocity constrained to

$$v_{i,j}^{\max} = (x_{i,j}^{\max} - x_{i,j}^{\min}) / n_r; \ j = 1, 2, 3, \dots, D$$
 (14)

A small n_r facilitates global exploration (searching new areas), while a larger one tends to facilitate local exploration (fine tuning of the current search area) and thereby attain an optimal solution in a lesser number of iterations.

4. CLUSTERING ALGORITHM

Feature extraction is done by applying standard statistical techniques to the contours of the modified S-matrix. These features have been found to be useful for detection, classification or quantification of relevant parameters of the time varying data streams. The five prominent statistical features are extracted from the S-transform output, as listed below:

1.
$$F_1 = \max(A) + \min(A) - \max(B) - \min(B)$$
 (15)

where A is the amplitude versus time graph from the S-matrix under disturbance and B is the amplitude versus time graph of the S-matrix without disturbance.

2. F_2 = Standard deviation of the magnitude versus time

spectrum obtained from ST matrix

$$= \left(\frac{1}{N}\sum_{j}\left(ST(j) - \frac{1}{N}\sum_{j}ST(j)\right)^{2}\right)^{1/2}$$
(16)

3. F_3 = Energy in the S-transform output = $|ST(n, j)|^2$ (17)

4.
$$F_4$$
 = Total harmonic distortion (THD) =

$$\sqrt{\sum_{n=2}^{N} \left[V_n \right]^2 / \left| V_1 \right|}$$
(18)

where N is the number of points in the FFT, V_n the value of the nth harmonic component of the FFT.

These features are used to classify the time series data into several classes by using a hybrid clustering technique based on PSO and K-means algorithms. The K-means algorithm tends to converge faster (after less function evaluations) than the PSO, but usually with a less accurate clustering. This section shows that the performance of the PSO clustering can further be improved by sending the initial swarm with the result of the K-means algorithm. The hybrid algorithm first executes K-means algorithm once. In this case the K-means clustering is terminated when the maximum number of iterations is exceeded, or when the average change in centroid vectors is less that that 0.0001 (a user specified parameter). The result of the K-means

algorithm is then used as one of the particles, while the rest of the swarm is initialized randomly.

Figs. 2(a) and 2(b) depict the plot of the features F_1 versus F_2 for both the K-means clustering and PSO variants, respectively. The figures clearly show the four number of clusters (Cluster-1 to Cluster-4) belonging to steady disturbance patterns like voltage sag (or dip), voltage swell (rise), harmonics and normal signal time series, and transient disturbance patterns like oscillatory transient, notches, and spikes. It can be seen that Cluster-1 belongs to voltage dip / sag database; Cluster 2 belongs to voltage rise / swell database; Cluster-3 to harmonics (both high and low), and Cluster-4 belongs to the transient signal time series database. From Figs. 2(a) and 2(b) it can be seen that there may be a number of misclassifications for K-means clustering, present in the data is <5% of the fundamental component. When the harmonic content is more than 7 to 10 % of the fundamental component, the cluster 6 becomes prominent and shows such data points.

5.FUZZY DECISION TREE FOR CLASSIFICATION

Decision Tree (DT) [15-17] is a data mining classification algorithm for high-dimension pattern classification. Each internal node of the tree tests the value of a predictor and consequence is represented by each branch of the tree. Leaf nodes represent the terminating node of the tree that classifies a particular event. Further, the confidence and support of a decision can be assigned to the terminating node. In a simple decision tree the number of the predictors is limited, while in a large dimensioned problem, DT software needs to be used. Here the DT analysis uses the significant features F_1 , F_2 , F_3 , and F_4 to provide the decision of a power network disturbance event. Although a simple binary decision tree can provide classification, an uncertainty can arise due to the overlapping nature of the features of different disturbance classes. This necessitates the construction of a fuzzy decision tree. The fuzzy DT rules are described below:

Rule 1: If $0 < F_1 < 0.95$ and $.01 < F_2 < .06$, Then Cluster-1 Rule 2: If $1.15 < F_1 < 1.8$ Then Cluster-2 Rules 3: If $.9 < F_1 < 1.1$ and $.01 < F_2 < .06$, Then Cluster-3 Rule 4: If $.85 < F_1 < 1.1$ and $.09 < F_2 < .18$, Then Cluster-4 Rule 5: If $.01 < F_2 < .03$ and $.02 < F_3 < .06$, Then Cluster-5 Rule 6: If $.02 < F_2 < .03$ and $.06 < F_3 < .15$, Then Cluster-6 Rule 7: If $.98 < F_1 < 1.06$ and $.005 < F_3 < .03$, Then Cluster-7 Rule 8: If $.9 < F_1 < .98$ and $.05 < F_3 < .09$, Then Cluster-8 Rule 9: If $1.0 < F_1 < 1.2$ and $.05 < F_3 < .1$, Then Cluster-9

Trapezoidal membership functions are used to represent the fuzzy sets described in the decision tree rule base.

Using the decision tree (Fig.5) the power disturbance events are classified depending on the belongingness of an event to a particular cluster. Accordingly, the following clusters are used to classify a given power disturbance event;

Cluster 1: voltage sag, Cluster 2: voltage swell, Cluster 3: harmonics, Cluster 4: Very short duration Transient event, Cluster-5: Low order harmonics, Cluster-6: High order harmonics, Cluster 7: Oscillatory transients, Cluster 8: Spikes, Cluster 9: Notches.

A total of 3000 signals are used for training and another 1000 separate signals are used for testing the clustering based classification. The average classification accuracy is found to be very nearly 100% for power disturbance databases without any noise in the data streams. However, when the databases are contaminated

with noise, the accuracy falls down marginally to an average of 99%. Table-3 provides DT based classification of a several power network disturbance time series databases in comparison to other existing classification techniques.

6. CONCLUSION

An evolutionary clustering technique has been developed by hybridizing the K-means algorithm and PSO for time-series clustering of power signal disturbance data. It can be considered as a viable and an efficient heuristic to find optimal or near-optimal solutions to clustering problems of allocating N data points to K clusters. The proposed method is very efficient and simple to implement for clustering analysis when the number of clusters is known a priori.

7. REFERENCES

- J. Han, M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, San Francisco, 2001 pp. 346–389.
- [2]. H. Ralambondrainy, "A conceptual version of the K-means algorithm," Pattern Recognit. Lett., vol. 16, no. 11, pp. 1147–1157, 1995.
- [3]. T. Zhang, "Data clustering for very large datasets plus applications," Ph. D. dissertation, Dept. Comput. Sci., Univ. Wisconsin, Madison, 1997.
- [4]. Z. Struzik and A. Sibes. Measuring time series similarity through large singular features revealed with wavelet transformation, In Proc. of the 10th Intl. Workshop on Database and Expert Systems Appl., pg. 162–166, 1999.
- [5]. T.W. Liao, "Clustering time series data a survey", Pattern Recognition, vol.38, 2005, pp.1857-1874.
- [6]. D. Buras et al, Wavelet and neural structure: A new tool for diagnostic of power system disturbances, IEEE Trans. on Industry Applications, 37(1) (2001) 184-190.
- [7]. W.Lin, C.Wu, Chia-Hung Lin, Fu-Sheng, Detection, classification of multiple power quality disturbances with wavelet multiclass SVM, IEEE Trans. on Power Delivery, 23(4)(2008)2575-2582.
- [8] A.M. Gouda, M.M.A. Salama, S.H. Kanoun, and A.Y. Chikhani, Pattern Recognition Applications for Power System Disturbance Classification, IEEE Trans. on Power Delivery, 17(3) (2002) 677-682.
- [9] Z.-L. Gaing, Wavelet-based neural network for power disturbance recognition and classification, IEEE Trans. on Power Delivery, 19(4) (2004) 1560–1568.
- [10] Haibo He and J. A. Starzyk, A Self-Organizing Learning Array System for Power Quality Classification Based on Wavelet Transform, IEEE Trans. on Power Delivery, 21(1), (2006) 286-295.
- [11] I.W.C. Lee, P.K.Dash, "S-transform based intelligent system for classification of power quality disturbance signals", IEEE Transactions of Industrial Electronics, vol.50, no.4, pp.800-806.
- [12] P, K Dash, B.K Panigrahi, and G.Panda "Power Quality Analysis Using S Transform", IEEE Transaction on Power Delivery, vol.18,No 2,April 2003.

- [13] Nicolaos B. Karayiannis, and James C. Bezdek, "An Integrated Approach to Fuzzy Learning Vector Quantization and Fuzzy C-means Clustering" IEEE Trans. on Fuzzy System, vol. 5, no. 4, November 1997.
- [14] R.A.Brown, Richard Frayne, "A Fast Discrete S-Transform for Biomedical Signal Processing", 30th Annual IEEE EMBS conference, Vancouver, Aug.2008, pp.2586-2589.
- [15] Selim, S. Z. and Ismail, M. A., "K-means Type Algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality," IEEE Trans. Pattern Anal. Mach. Intell. vol. 6, pp.81-87, 1984.
- [16] Kennedy, J. and Eberhart, R., "Particle Swarm Optimization," Proc. of IEEE International Conference on Neural Networks (ICNN), Perth, Australia, vol.4, pp.1942-1948, 1995.
- [17] Eberhart, R. and Kennedy, J., "A New Optimizer Using Particle Swarm Theory," Proc. 6th Int. Symposium on Micro Machine and Human Science, pp. 39-43, 1995.



Fig.1 Time series clustering using feature based approach.



Fig.2 Fuzzy set for Decision Tree



Fig. 3 F₁- F₂ K-Means PSO Cluster 1 Sag, Cluster 2: Swell, Cluster 3: Normal + harmonic,



Fig.4 F₁ - F₂ Fitness Plots



Power Signal	F ₁	F ₂	F ₃	F ₄
Disturbances				
Normal	1.002	0	0.046	0
Sag (60%)	0.597	0.0536	.0315	.0318
Swell (50%)	1.50	0.0129	.076	.015
Momentary	0.0724	0.035	.019	.0350
Interruption				
(MI) (5%)				
Harmonics	1.0	0.0339	.0556	.141
$(0\% 3^{rd} +$				
10% 5 th)				
Sag with	0.601	.0228	.0408	.1139
Harmonic				
(60%)				
Swell with	1.5	.0219	.079	.1155
Harmonic				
(50%)				
Flicker (5 Hz,	0.987	.0168	.026	.0186
4%)				
Notch +	0.939	0.131	.0529	.136
harmonics				
Spike +	1.065	0.141	.0627	.1308
harmonics				
Transient (low	0.493	0.138	.0163	.01
frequency)				
Transient	1.052	0.149	.014	.043
(high				
frequency)				

Table-1: Features extracted from modified Fast modified S-transform