Development of Cluster Algorithm for Grid Health Monitoring

R.K. Pandey, *Senior Member IEEE*, Satish Kumar and Chandan Kumar

***Abstract*—This paper describes the use of K-means clustering algorithm to mine the Synchrophasor data from PMUs. PMUs are newly developed tools for monitoring the grid health by measuring grid parameters such as voltage, current, frequency, rate of change of frequency and phase angle with high sample rate and time stamping. The large amount of data produced by PMUs can help the grid operator in the stable operation of the grid. But this data cannot be useful until it is mined properly using different methods. Application of K-means clustering algorithm has been done to extract important information from the Synchrophasor data. This information helps the operator to take real time decisions and ensures the grid stability.**

Keywords—: Cluster, Phasor Measurement Units (PMUs), Rate of Change of Frequency (ROCOF) , SCADA, Wide Area Monitoring System (WAMS).

I. INTRODUCTION

The complexity and challenges in the power system operation are increasing with the growing size of power system network. In such a vast interconnected system, real time monitoring has become a major concern for a stable operation of the power system. SCADA has been incapable of solving the rising challenges in the power network. In new framework, PMU provides GPS time stamped Synchrophasor data which has been proved promising for precise measurements of grid parameters and also to detect even the minute changes in a dynamically changing system that was not possible until now.

The phasor data transmitted by the PMUs are time tagged with the UTC time and called as Synchrophasors. Each PMU sends signals of interest every second in IEEE C37.118-2005 or 2011 to a Phasor Data Concentrator (PDC) which arranges the data according to their time stamp and finally collected by a super data concentrator. Since all the PMUs are synchronized with the UTC and data is being updated at a very high rate, this data can be used for monitoring grid health in real time. The measurements recorded by PMUs help the operator to identify the problems of power network in real time if any. With this multi-dimensional high resolution data, even minute changes in the grid can be visualized by the grid operators which were not possible with SCADA earlier. PMUs monitor the phase angle difference of various nodes and allow the power control decision in better way. Also, it enables visualization of voltage and current phasors, frequency, rate of change of frequency and angular separation at every few millisecond interval (in India at 40 milliseconds) in the Load Despatch Centre [1-2]. Thus the dynamical behavior of power system can be observed in almost near real-time at the control centre which hitherto was possible only in offline mode in the form of substation Disturbance Records (DR) or through offline dynamic simulations performed on network model. Hence a paradigm shift from state estimation to state determination can be easily achieved, which hitherto was not possible. So, the limitations with the conventional SCADA system are successfully addressed by the PMUs.

It is well known that the data collected by PMUs comes under big data. The rate of generation of this data is about 400-1000 observations per second, resulting in large volumes of phasor data. As a greater number of PMUs is installed for better observability, the growth in the size of the data occurs exponentially. In wide-area monitoring systems (WAMS) containing a large number of PMUs, the amount of data generated may be so large that it can be very difficult to handle this data for effective control decision. For example, the Tennessee Valley Authority (TVA) presently handles 120 online PMUs with 3.6 billion measurements archived per day with a storage size of 36GB [3].While these data will help an utility to understand its regional/national grid operations at a granular level and enable more targeted operational decisions, ultimately leading to more efficient grid operations, managing the data flow and its analysis is very tedious and time consuming, if viewed classically.

Since this data contains a large amount of information which can be useful for the normal operation of grid, it becomes imperative to process this data. However, processing of such a large volume of data is a big challenge as on date. Such data cannot be mined using simple techniques because it needs to be processed efficiently without losing any important information and also this cannot be stored owing to its size. So, an algorithm that processes entire data is required. The algorithm should be competent enough with the rate of arrival of the data else the ultimate goal of installing PMUs is lost. Data mining techniques combined with statistical analysis can be used to extract important information from this data. In this paper, K-means clustering algorithm has been discussed which can be used for analysis of Synchrophasor data. The algorithm has been tested on MATLAB and experimental results have been discussed.

Section II provides the background for data mining and section III discusses the K-means clustering algorithm. The analytical studies have been discussed in section IV while section V concludes the work.

II. DATA MINING CONCEPT

It is the process of extracting useful information from large data sets. The fundamental idea is to discover patterns in the

---

R. K. Pandey, Satish Kumar are with Electrical Engineering Dept., Indian Institute of Technology (Banaras Hindu University), Varanasi, India (e-mail: rpsneh@yahoo.co.in, satish.kumar.eee11@iitbhu.ac.in).
Chandan Kumar is with Power Grid Corporation of India Limited (PGCIL). (E-mail: chandan8240000@gmail.com).

data and the identification of unusual data records that might be useful or that contain errors. It involves the association of different variables. It may also involve the clustering into groups of “similar” objects or use of functions to model available data with the least error.

Data mining has been applied to several power system applications [4-7]. While data mining techniques have been applied in different power system applications, these methods have achieved new interest in the context of Smart Grids since integration of data and information systems is one of the key advantages of the Smart Grid [8]. Over the last few years, there is an increasing trend of deploying several hundred Phasor Measurement Units (PMUs) in the various grids of the world to monitor the grid health accurately, for example in North American grid, PMUs are deployed and they aggregate their data into Phasor Data Concentrators (PDC). These PDCs stream their data to a super PDC at the Tennessee Valley Authority (TVA). To cover the grid adequately, it is estimated that at least one third of the bulk power systems locations should be monitored by PMUs, thus requiring about thousands of PMUs and other resources for processing billions of data samples per day. Hence, the most important work is to handle the increasing amounts of data acquired by the network of new sensors and measurement devices. As volumes of such data are becoming gigantic, terms such as “data explosion” or “big data” are being used more frequently to describe the situation of power systems with respect to data. All of this data could be utilized for improved performance of the power system. However, this will require proper mining and right interpretation. Analytical methods such as based on the advanced ideas of statistical tools, pattern recognition and intelligent controls will increasingly become the need of the hour.

Hence, there is a critical requirement for development of highly efficient algorithms that could uncover important patterns and information in power system data and take advantage of what the Smart Grid offers. In order to address this rising problem of data explosion, data mining comes up as a natural choice.

Cluster analysis is a major part of data mining. The ultimate aim of clustering algorithms is to group objects based on the information that describes them. Clusters can be considered as classes, to which the clustering process automatically assigns objects. Clustering is considered as an unsupervised learning since it does not require class labels to be known ahead of time. Several clustering algorithms are available in data mining literature [9]. Figure 1 shows few data points and one possible way to divide them in clusters on the basis of their colour. It is important to note that there can be multiple ways to cluster objects since the same set of objects can be classified differently according to different attributes. So selection of the appropriate attribute and clustering algorithm for any application is a matter of great significance.

Clustering is a powerful tool as it can identify groups of similar attributes or similar patterns. If there is any unusual pattern or object, it will usually form a single-object cluster. Such unusual behavior can be used for spotting “interesting” objects and patterns.

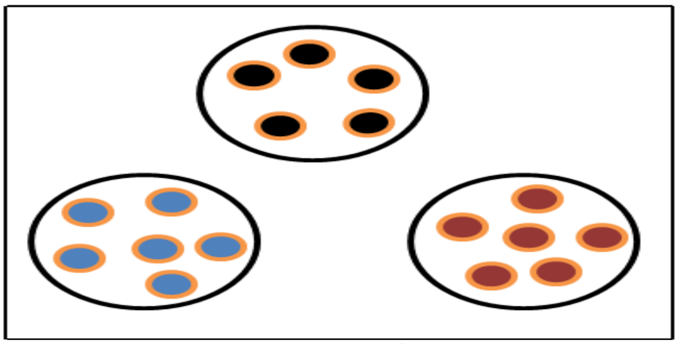
Figure 1: Demonstration of Clustering

III. K-MEANS CLUSTERING ALGORITHM

K-means clustering is an unsupervised learning algorithm for classifying a given data set into a certain K number of clusters. Following are the steps for K-means clustering algorithm:

Step1 – Set the value of K.
Step 2 - Define K centroids, one for each cluster. Centroids from the given data set should be selected in a way that they are as much as possible far away from each other.
Step 3 - Take each point belonging to the given data set and associate it to the nearest centroid forming clusters. When all the points have been grouped, the first iteration is completed.
Step 4 - Now K new centroids are recalculated as centers of the clusters resulting from the previous step.
Step 5 - Once again the same data set points are grouped so that each one is now associated with its nearest new centroid.
Step 6 - This process continues till there is no change in the location of the centroids. In other words centroids do not move any more.

IV. EXPERIMENTAL RESULTS

The analysis with K means Clustering Algorithm has been done using data sets taken from utility- POWERGRID for different conditions and oscillations in the grid and the corresponding plots for one PMU at location A in WRLDC are presented here. The value of K was assigned equal to 10. Thus 10 clusters were generated and corresponding plots for the clusters are shown below. In cluster 1, it can be noted that minimum frequency occurs at 8:38:54:200 and corresponding value is 49.805424 hertz, which is below the normal range of operation. Also, there is large dip in the voltage of phase c as compared to the other phases. In cluster 2, it can be noted that there is further dip in the frequency and phase voltages. Similarly, this can be explained for other clusters as well. From all the clusters (Figs.2-21), it is noted that minimum frequency occurs in cluster 8 and it is 49.701287 hertz at time stamp equal to 8:38:46:360. Similarly, the minimum voltage occurs in cluster 5. So, by analyzing the below clusters, proper control action can be initiated at the appropriate time and proper functioning of the power grid can be ensured. It may be noted that the smart control may be evolved with real time cluster and adequate controllers may be switched into the network for improved operational features of power utility even under varying system conditions.

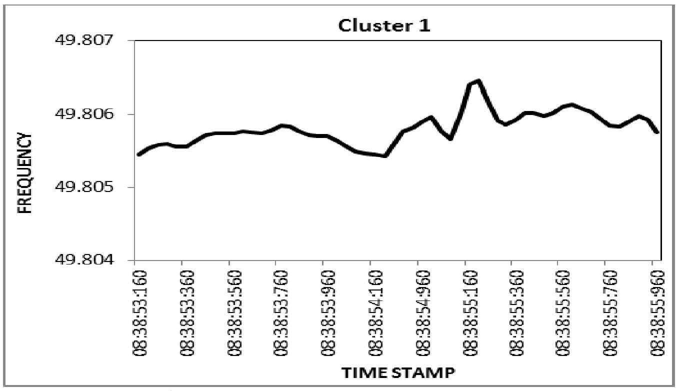

Figure 2: Cluster 1 for PMU signal record at A

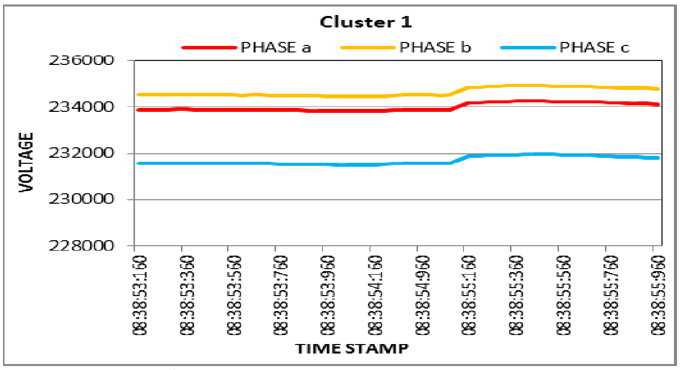

Figure 3: Cluster 1 for PMU signal record at A

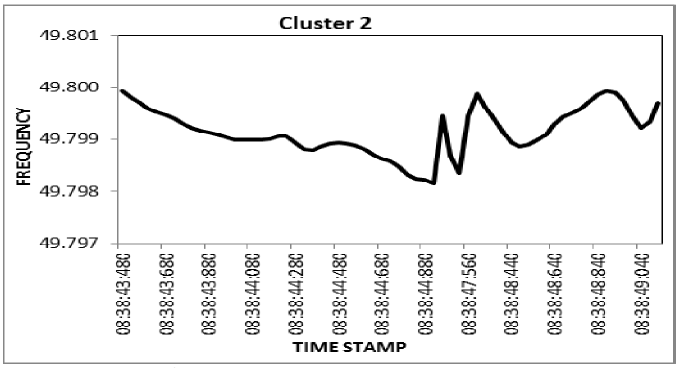

Figure 4: Cluster 2 for PMU signal record at A

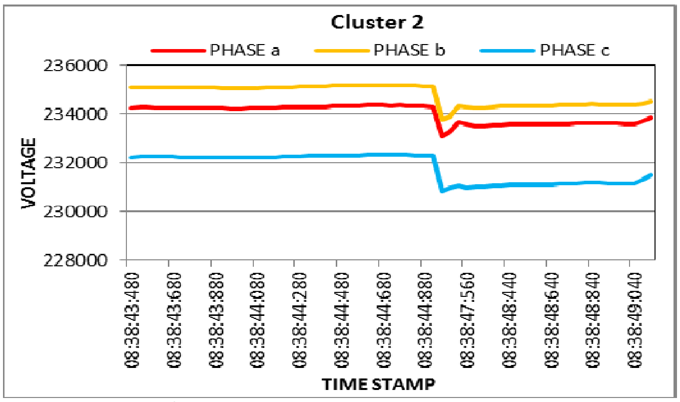

Figure 5: Cluster 2 for PMU signal record at A

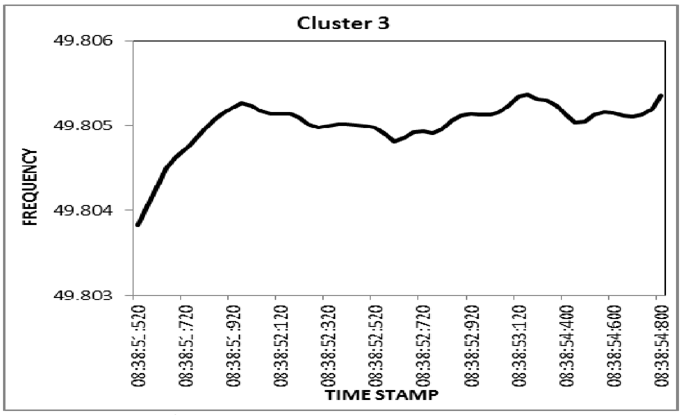

Figure 6: Cluster 3 for PMU signal record at A

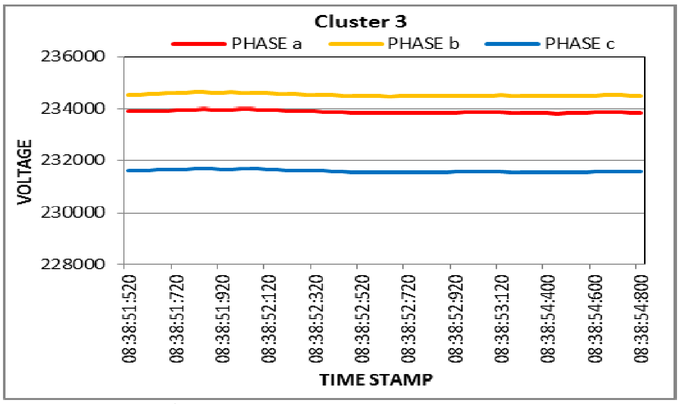

Figure 7: Cluster 3 for PMU signal record at A

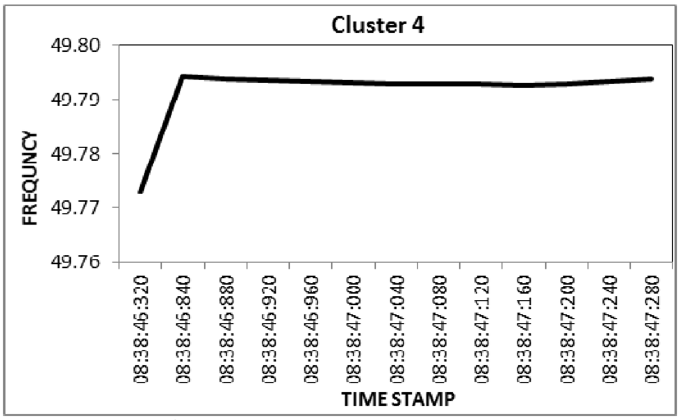

Figure 8: Cluster 4 for PMU signal record at A

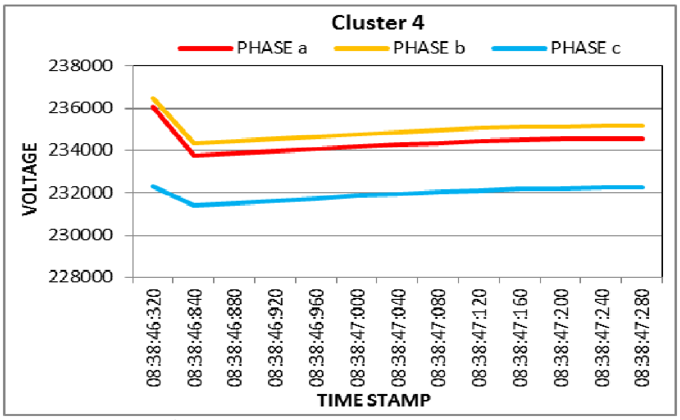

Figure 9: Cluster 4 for PMU signal record at A

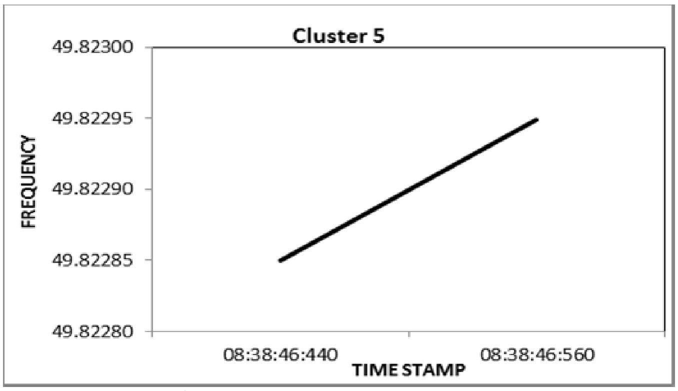

Figure 10: Cluster 5 for PMU signal record at A

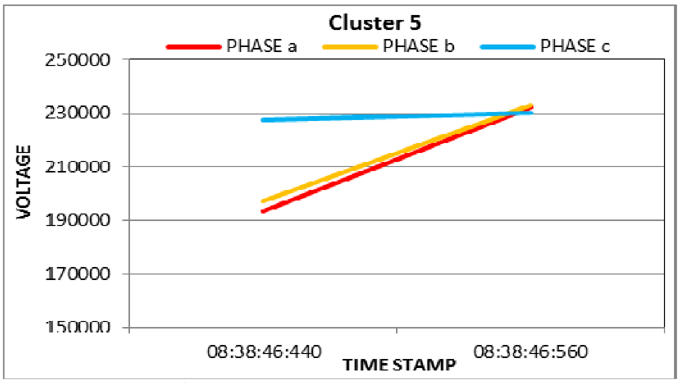

Figure 11: Cluster 5 for PMU signal record at A

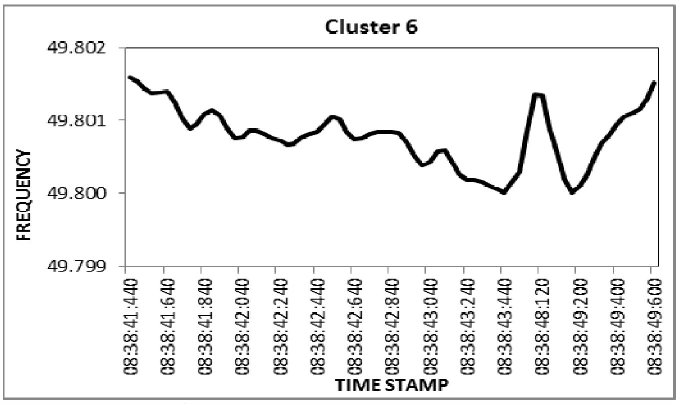

Figure 12: Cluster 6 for PMU signal record at A

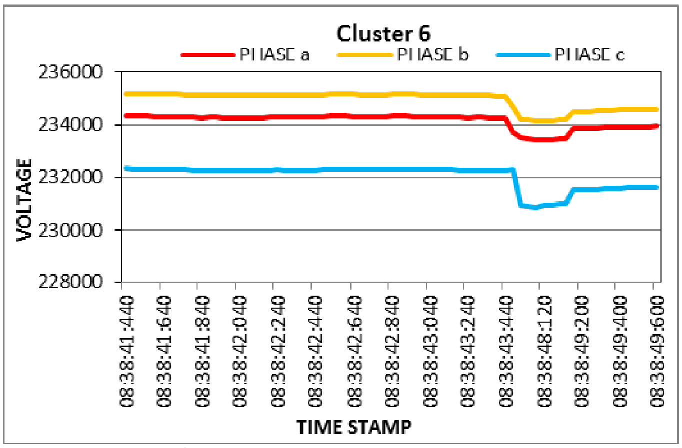

Figure 13: Cluster 6 for PMU signal record at A

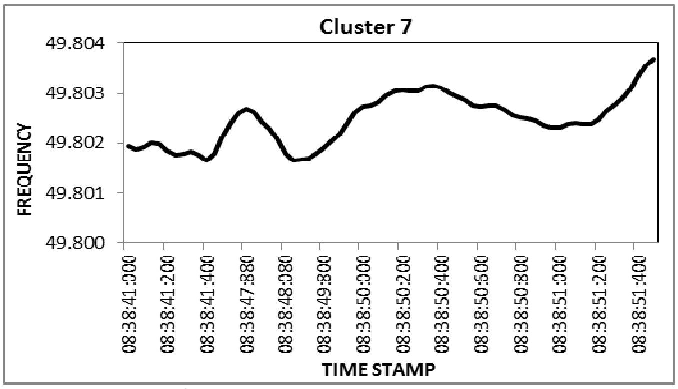

Figure 14: Cluster 7 for PMU signal record at A

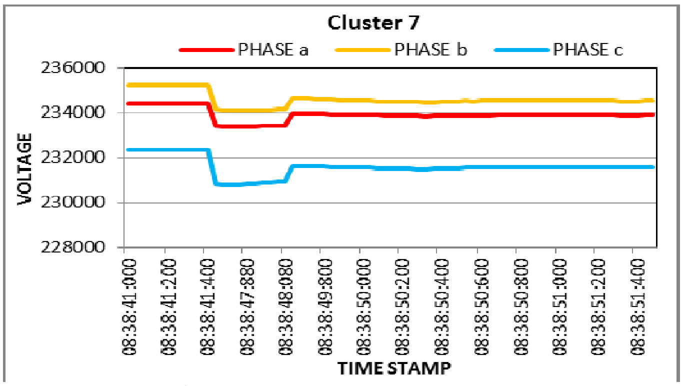

Figure 15: Cluster 7 for PMU signal record at A

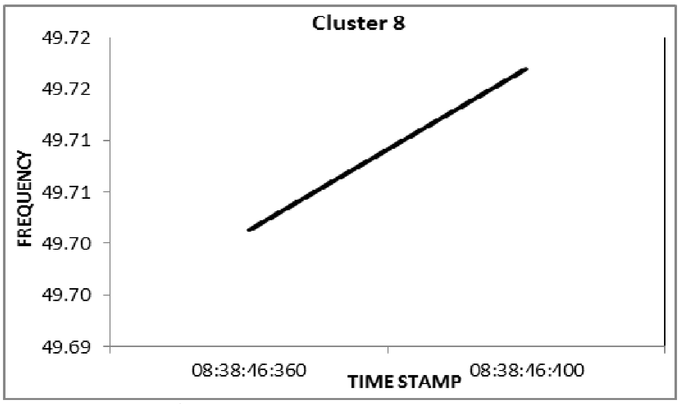

Figure16: Cluster 8 for PMU signal record at A

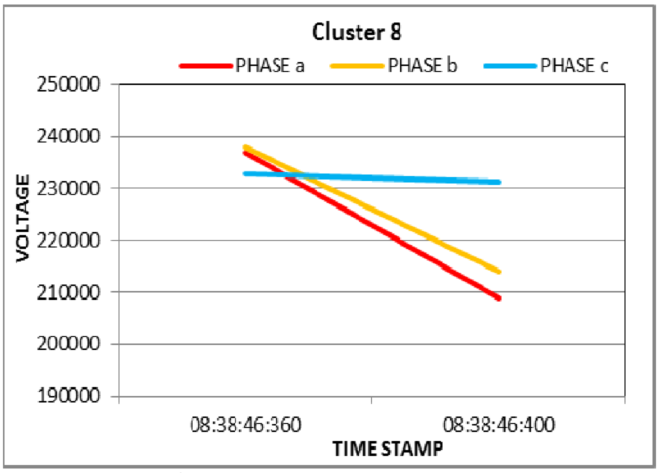

Figure 17: Cluster 8 for PMU signal record at A

Figure 18: Cluster 9 for PMU signal record at A

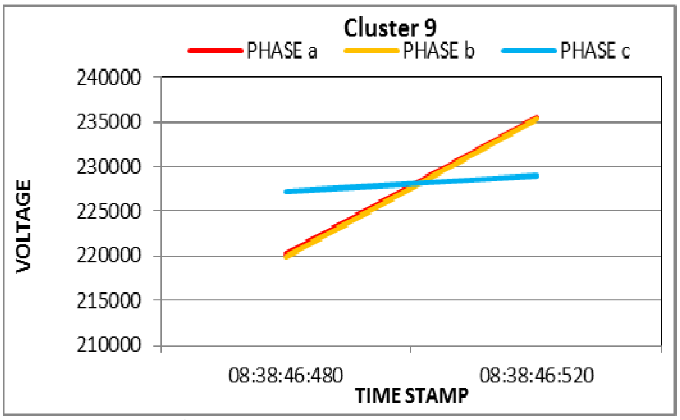

Figure 19: Cluster 9 for PMU signal record at A

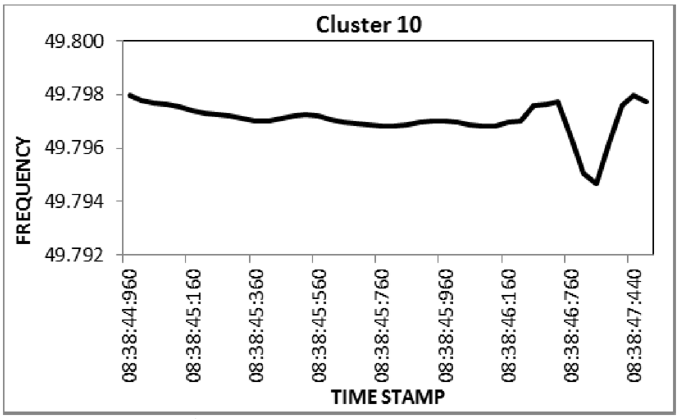

Figure 20: Cluster 10 for PMU signal record at A

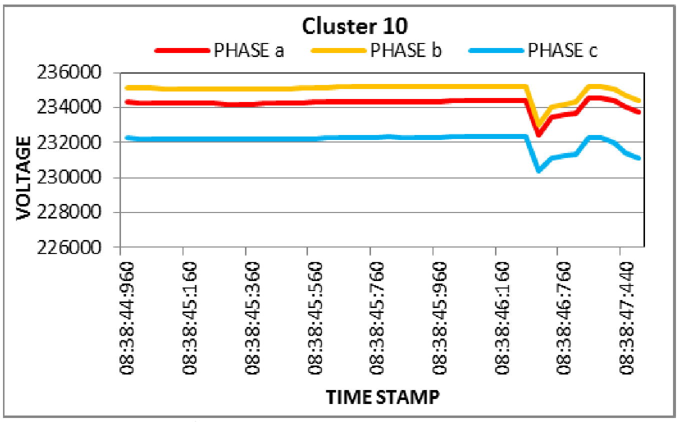

Figure 21: Cluster 10 for PMU signal record at A

V. CONCLUSION

This paper presents an application of clustering algorithm for event detection and related control injection in power network. The methodology of K-means clustering scheme has been given which can be implemented to extract important information from the large datasets. The control signals can be initiated depending on the severity of the fault and also specific controller may be identified for improved Grid behaviour.

The deviation, in any of the quantity being measured, by PMUs can be suitably detected and control action can be initiated almost in real time. The algorithm can be generalized by developing techniques in which the lower and upper threshold of the data can be detected by the algorithm itself in learning mode without the operator manual intervention in setting the threshold limits. This algorithm can be applied over the data received from PMU as signals, i.e., rate of change of frequency, voltage, frequency, current, angle with an appropriate choice of clusters. Further extension of the K-means clustering algorithm can be done by applying it on the real time data in windowing frame concept.

VI. Acknowledgment

The authors acknowledge with thanks the guidance and support given by managements of POSOCO for permitting the publication of this paper. The authors are also thankful to WRLDC personnel for their support. The views expressed in this paper are of authors and not necessarily that of the organizations they represent.

VII. References

[1] R. K. Pandey, P. Pentayya, A. Gartia, Rajkumar A. and Chandan Kumar “Operational Reliability Enhancement with PMUs in Indian Power Network”, IEEE CATCON, Kolakata, 2013.

[2] “Synchorphasors Initiative in India” POSOCO, India, June 2012

[3] “New Technology can Improve Electric Power System Efficiency and Reliability”, Today in Energy, U.S Energy Information Administration (eia), March 30, 2012.

[4] M.Sherwood, D. Hu and V.Venkatasubramanian, “Real-Time Detection of Angle Instability using Synchrophasors and Action Principle”, Bulk Power System Dynamics and Control-VII, August 2007, Charleston SC.

[5] G. Liu, J. Quintero, and V. Venkatasubramanian, “Oscillation Monitoring System based on wide area synchrophasors in power systems,” Proc. IREP symposium 2007. Bulk Power System Dynamics and Control - VII, August 19-24, 2007, Charleston, South Carolina, USA.

[6] Kathleen Wolf Davis, “On the NE Blackout’s Tenth Anniversary: PMU Data Analytics in Future Grid Reliability”, Intelligent Utility, Aug 13, 2013.

[7] PMU Enhanced Power System Operations, Interactive Energy Roadmap to Achieve Energy Delivery Systems Cyber security, Sept 2011.

[8]Chakrabarti S, Kyriakides E, “Optimal Placement of Phasor Measurement Units for Power System Observability”, IEEE Transactions on Power Systems, Vol.23, No. 3, pp1433-1440.

[9] A. Asheibi, D. Stirling, and D. Robinson, “Identification of load power quality characteristics using data mining,” in *Proceedings of IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, Ottawa, pp. 157-162, May 2006.