Polar Based Data Analytics for Load Profiling of Customer in Smart Grid

Amit Tiwari

Senior Project Associate

Indian Institute of Technology, Kanpur, India
amittwr00@gmail.com

Abstract: This paper proposes, load profiling of various users by using a polar projection technique on energy meter data (Kwh), resulting in reduction of computational complexity. K-means agglomerative clustering is used with a new distance metric(Polar distance) to find the optimum cluster for each load profile. To increase the speed of clustering, the stochastic-based algorithm is used, which uses only a part of data while training, taking lesser computational resources. We show that the polar-based stochastic approach not only reduces the computational complexity, but the resulting clusters are also more informative.

Introduction: AMI (Advanced Metering Infrastructure) is the collective term to describe the whole infrastructure from Smart Meter for two way-communication and all the applications that enable the gathering and transfer of energy usage information in real-time in the control center. AMI makes two-way communications with customers and is the backbone of the smart grid. The objectives of AMI can be remote meter reading for error-free data, network problem identification, forecast, demand response, load profiling, and energy audit.

If we deployed a large number of smart meter data can generate a massive amount of data. For example, if we deployed the thirty million smart meters and take data every half hour (30min), the generated data is 1.57 x 10^{13} in just one year. So, the characteristics of smart meter data are big data, which is a large volume and high heterogeneity. It becomes crucial to reduce the data dimensionality for real-time processing. Different supervised ,unsupervised and semi-supervised techniques are used for reduction in data dimensionality. for example kernel-based PCA, ICA and so on.

The reason for this new methodology is complexity involved while incorporating PCA,ICA etc.

According to this paper, this is the first time this approach has been used for data dimensionality reduction in big data analytics for reducing the search space in polar-based clustering, used the stochastic-based algorithm.

In this paper, we have used the unsupervised learning method to reduce the dimensionality of the meter data.

General Method for clustering: Generally used the Principal Component Analysis (PCA) but the complexity of PCA is $O\left(NM^2+M^3\right)$ and according to this paper computational complexity can be reduced to O(4NM) by using polar projection, where N is a number of samples and M is the dimension of the data.

Clustering: Total consumption and peak consumption are two important features for load profiling**.** Following polar projection the feature extraction and feature selection for clustering are performed simultaneously. Without polar projection, and polar distance metric the cluster obtained using K-means(euclidean distance) are classified only on the basis of their total consumption.

Polar projection for reduce dimension: The below figure shows, how polar projection is used to reduce the dimensionality of our original data

where, N= number of users, M= time stamp of meter data.

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \\ \vdots \\ X_N \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,m} & \cdots & x_{1,M} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,m} & \cdots & x_{2,M} \\ \vdots & \vdots & & & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,m} & \cdots & x_{n,M} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N,1} & x_{N,2} & \cdots & x_{N,m} & \cdots & x_{N,M} \end{pmatrix}$$

Let us have a , Projection vector as

$$P = \left[p_{1,} p_{1,} p_{1,}\ldots, p_{m,}\ldots, p_M\right]^T$$

Using P matrix, transfer X matrix(Fig 1) into Y

$$Y = XP$$

$$p_m = e^{i\theta_m} \qquad \theta_m = \frac{\pi(m-1)}{2(M-1)}$$

Where M= Total no. of sample, m= sample no. out of ‘M’.

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \\ \vdots \\ Y_N \end{pmatrix} = \begin{pmatrix} X_1 P \\ X_2 P \\ \vdots \\ X_n P \\ \vdots \\ X_N P \end{pmatrix} = \begin{pmatrix} \sum_{m=1}^{M} x_{1,m} p_m \\ \sum_{m=1}^{M} x_{2,m} p_m \\ \vdots \\ \sum_{m=1}^{M} x_{n,m} p_m \\ \vdots \\ \sum_{m=1}^{M} x_{N,m} p_m \end{pmatrix}$$

$$B = \begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \\ \vdots \\ B_N \end{pmatrix} = \begin{pmatrix} |Y_1| \angle Y_1 \\ |Y_2| \angle Y_2 \\ \vdots \\ |Y_n| \angle Y_n \\ \vdots \\ |Y_N| \angle Y_N \end{pmatrix} = \begin{pmatrix} r_1 \theta_1 \\ r_2 \theta_2 \\ \vdots \\ r_n \theta_n \\ \vdots \\ r_N \theta_N \end{pmatrix}$$

This results in two dimensional data from M dimensions. The polar angle contains the information about peak load of profile

[θ=f(time or m)] and magnitude ‘r’ contains information about total consumption of the profile.

Load profile clustering: This paper introduces a new distance metric for consumer loads.

Let $\varepsilon_k = [r_k, \theta_k]$ represent the polar coordinate of cluster center k. New distance metrics defines as

$$D(\in_k, B_n) = \sqrt{\left(\frac{r_n - r_k}{r_k}\right)^2 + (\theta_n - \theta_k)^2}$$

Where $B_n = [r_k, \theta_k]$

Euclidean distance is define as:

$$d(p,q) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$$

Where p, q = two point in Euclidean n-space

Identifying the appropriate number of clusters in K-means is a very important task. A large number of clusters complicates the clustering algorithm, whereas a small number can result in information loss. The sum of squared errors (SSE) technique is the most commonly used for this purpose, and it is used here . It is provided by

$$SSE = \frac{\sum_{k=1}^{K} \sum_{n=1}^{N} D(\varepsilon_k, B_n)}{N}$$

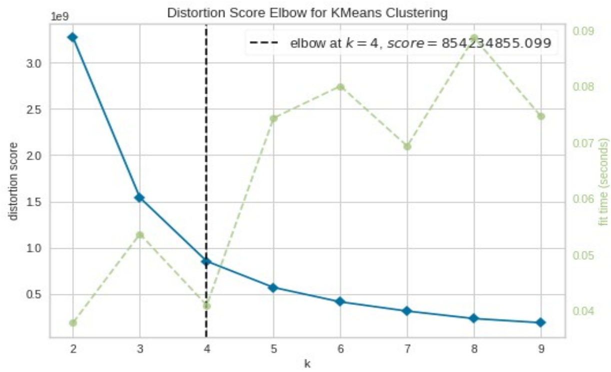

Figure 1: Optimum clusters=4 using elbow curve

Performance of the distance metric:

In order to compare the performance of distance metric K-means was first employed on the real part of polar data (r_n), using euclidean distance metric.

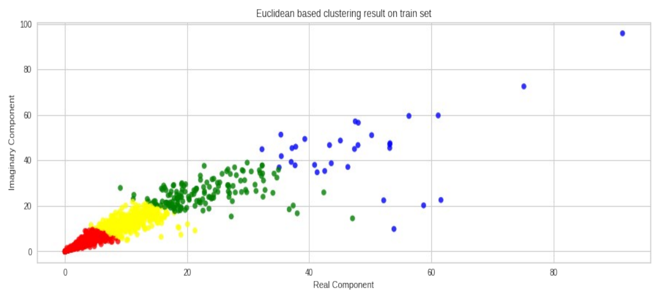

Figure 2: K-means clustering using Euclidean metric,where consumers are clasified according to total consumption(r_n)

The clusters formed are shown in Figure 3. Afterwards K-means was used with new polar distance metric on polar data (r_n , θ_n). The result is shown in figure 4:

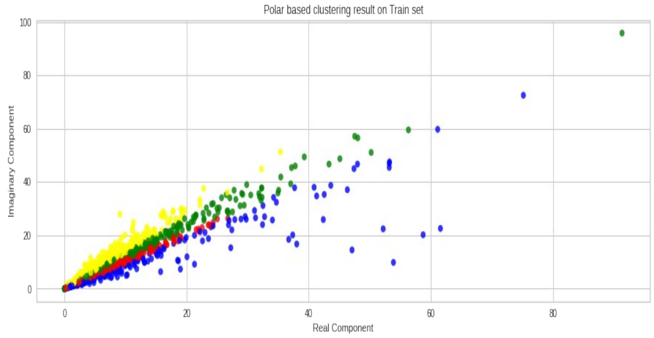

Figure 3: K-means clustering using Polar metric,where consumers are clasified according to total and peak consumption (r_n , θ_n)

Consumer Load Profiling:

The average consumption within each cluster, after obtaining the clusters using euclidean metric is shown in below figure:

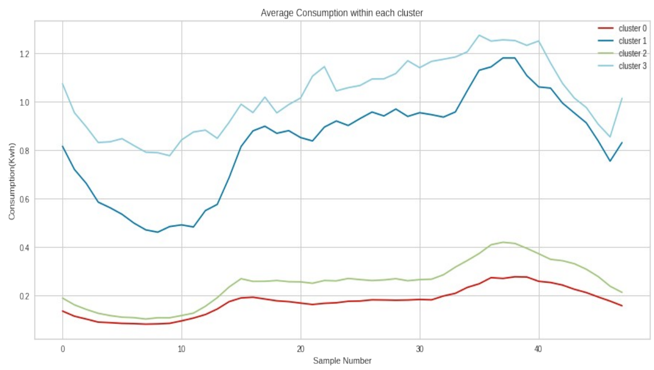

Figure 4: The average consumption within each cluster, after obtaining the clusters using euclidean metric

While the average consumption within each cluster, after obtaining the clusters using polar metric is shown in below figure:

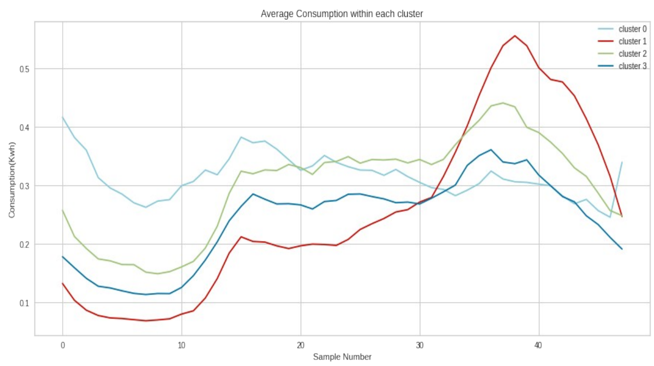

Figure 5: The average Consumption within each cluster, after obtaining the clusters using polar metric

Comparision:

Euclidean Clusters: If we average the load of all users in particular clusters, we can see that the cluster with average of user peak load and user total load i.e figure 6 and figure 7 shows two different clusters which also means that the peak user might lie in either of these clusters.

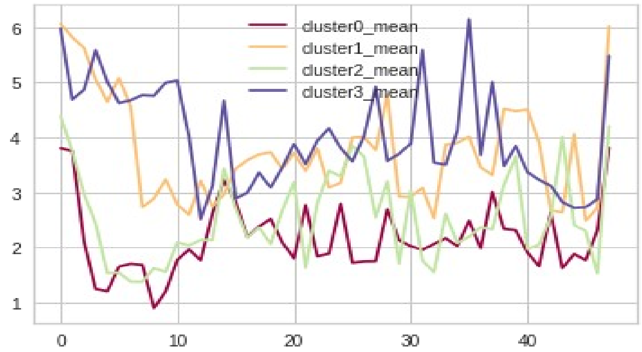

Figure 6: Mean of all “user Peak load” lying in particular cluster

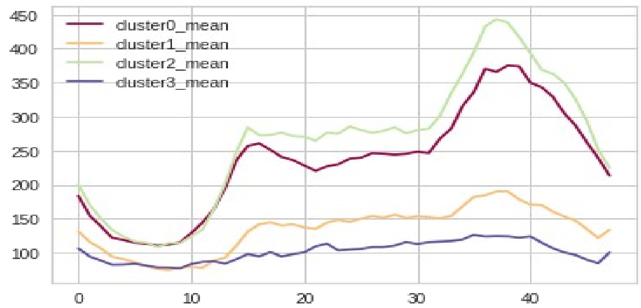

Figure 7: Mean of all “user Total load” lying in particular cluster

Polar Based Clusters: On the other hand, if we average the load of all users in polar clusters, we can see that the cluster with average of user peak load and user total load i.e figure 8 and figure 9 shows same cluster with high load profile.

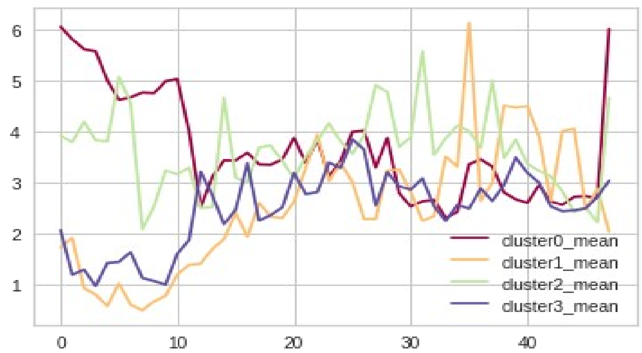

Figure 8: Mean of all "user peak load" lying in each cluster

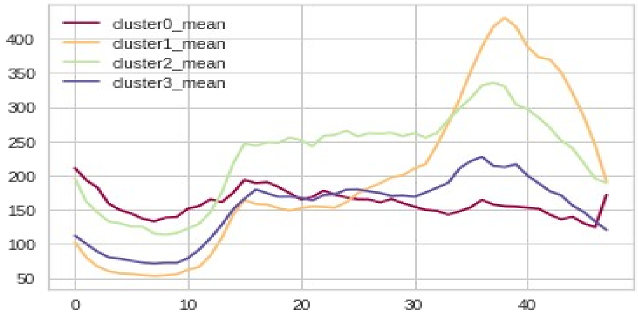

Figure 9: Mean of all "user Total load" lying in each cluster

This means our clusters in polar based k-means(custom polar dsitance metric) do account for peak and total load of customers.

Conclusion:

A polar-based dimensionality reduction technique was proposed in this article to reduce the complexity associated with clustering large smart meter datasets. A new distance metric based on total and peak consumption from consumer load profiles was used along with a stochastic approach which utilises a subset of dataset in order to train the model.

The load profiles obtained in figure 4, are obtained using euclidean distance and only total consumption of profile and therefore they do not tell the peak load of cluster but average total-consumption only. This means that the graph of each clusters remains within a average load range, of that cluster and can not tell about the peak hours for users of that cluster.

While the profiles obtained in figure 5 are more informative and can show the peak load of overall cluster and is the result of using polar projection as proposed in this paper.The cluster profile will not be bounded by a range and thus a better demand and response sytem can be incorporated further to utilise the result of this methodology. e.g Smart Billing system using load profiling etc.