CLUSTER ASSOCIATION USING SUSTAINABILITY TO LOAD PROFILE DETERMINATION

Kalyani.K[1] ,Chakravarthy.T[2]

[1]T.U.K.Arts College, Karanthai, Thanjavur, 613 002, India.

[2]AVVM Sri Pushpam College, Poondi, Thanjavur, 613 503, India.

Abstract: *The chief aim of this paper is to develop an effective approach to the issue of load profile clustering by applying Cluster Association using Sustainable Energy based Bees Optimization (CASEBO) algorithm. While, intelligent metering solutions like Automated Meter Reading (AMR), Automated Meter Infrastructure (AMI) are in place to address the current issues prevailing in the domain of electricity markets, algorithm using Improved Replacement In Best Optimization has been proved beneficial and uncomplicated to apply within a selective database. In this study Load Profile (LP) clustering distribution networks based on the shape of the load profile was studied for fitness function in the selected LP clustering. The results clearly indicate that LP clustering has advantages in providing metering solutions to consumers who do not possess digital metering which can be easily operated with trivial changes in the calibrations.*

Key words:- *Load profiling, honey bee modeling , Cluster Association using Sustainable Energy based bees Optimization(CASEBO)Algorithm, clustering techniques.*

I. INTRODUCTION:

The contemporary scenario of electricity market is flowing with a range of distribution marketing solutions like AMR or AMI systems in order to address the challenges prevailing in metering technologies in general and in particular a customer without digital technology. In pursuit of this issue, LPing techniques are extensively adopted to facilitate customers access the retail market and for tariff development purposes. Perhaps, this technology categorizes customers based on the shape of the year load profiles and generates Typical Load Profile (TLP) that can be used to formulate model load from distribution system.In this, customer without digital meter is assigned a consumer category so as to get a unique profile and behavior as an outcome of the specified TLP to the corresponding category. A broad range of methods have been proposed and tested on different load profile databases, such as K-means or hierarchical clustering, self-organizing [1, 2, 3]. The present scenario of electricity market is rolling with a range of intelligent maps, neural networks, fuzzy systems, statistical methods or more recently, the support Vector Clustering approach [4, 5,6,7]. This study proposes a new approach to the LP clustering by applying Cluster Association using Sustainable Energy based bee'salgorithm (CASEBO). Furthermore, due to the robustness and originality of this method.

There are significant benefits like product quality of the results with effortless and trivial change on certain simple parameters. Indeed, it has greater efficiency than alternative clustering approaches. This paper highlights briefly the overview of the most popular clustering techniques. Although data clustering aims to find structures in heterogeneous collection of data, these structures describe groups of data within a similar inside a group and dissimilar between different groups. The end result of the clustering algorithm or methodology depends mostly on the classification criterion and to separate similar and dissimilar data. Due to the advancement of technology in this field, lot of changes occurred during the last two decades which enabled to offer solution by introducing software based algorithm and improving the existing infrastructure facilities.

II. LOAD PROFILE CLUSTERING:

Load Profile is a broad term that can refer to a number of different forms of data. It can refer to demand and consumption data or it can be a reference to derived data types, such as Regression and Profile Coefficients. However, all these data types have one thing in common that they represent the pattern of electricity usage of a segment of supply market customers. A load profile gives the Half-Hourly (Settlement Period) pattern or 'shape' of usage across a day (Settlement Day), and the pattern across the Settlement year, for the average customer of each of the eight profile classes. It is the proportion of demand in each settlement period that is of interest to the Settlement System[8,9].

Cluster exploration is a term used to designate a family of statistical measures specifically designed to notice classifications within complex data sets. The aim of cluster analysis is to bunch objects into clusters so that objects within one cluster share more in common with one another than they do with the objects of other clusters. Consequently, thepurpose of the analysis is to arrange objects into relatively similar groups based on multivariate observations. The objective of Clustering data is to capture the structure in a heterogeneous group of data. These hierarchies define collections (or) clusters of data which are unambiguous inside a group and ambiguous between different groups. The end solution of a clustering algorithm or hierarchy depends in great extent on the classification used to partition of unambiguous and ambiguous data. While, investigators in the behavioral and social sciences are often interested in clustering people, clustering nonhuman objects is common in other disciplines [10].

Thereby, clustering algorithm is used to determine a load profile type and to analyze the demand load in a distribution substation. It is also important to understand the difference between clustering (unsupervised classification) and discriminate analysis (supervised classification). In supervised classification, a collection of labelled (reclassified) patterns are provided. Nevertheless, the issue is to label a newly encountered, yet unlabelled, pattern. Perhaps, the given labelled (training) patterns are used to study the descriptions of classes that in turn are used to label a new pattern. In the case of clustering, the struggle is to group a given collection of unlabelled patterns into meaningful clusters. In a sense, labels are associated with clusters also, but these category labels are data driven that is, they are obtained solely from the data. For instances, clustering algorithms can be applied in many fields such as Web optimization, Finance, image processing, Marketing, Biology, Libraries, Insurances, City planning, Earthquake studies etc.,[11].

III. SIGNIFICANCE OF LOAD PROFILE CLUSTERING (PROBLEM DEFINITION)

While, Clustering algorithms are applied in different kinds of applications such as web optimization, finance, biology, image processing etc. The clustering problem defined in this paper refers to the load description based on electricity distribution network and one of the consumer models is Typical Load Profile (TLP).

A TLP explain the hourly values of electricity consumption on a daily basis and associated to consumer category. TLPs can be delineated for residential, commercial, industrial and for seasonal factors. It can be established for seasonal factors of climatic conditions that are likely to occur in future using regressed technique. Most widely used approach to structure TLP consist of gathering actual LP for various consumer categories, metered in network supply points and process them using clustering algorithm to build TLPs. In order to setup a TLP portfolio for any public utility, it must define a set of TLP that can accurately provide load characteristics for all consumers in its self as a network. If the portfolio includes the maximum TLP and extensive consumers, then It is the best representation of consumers in terms of accuracy.

The customer load profile clustering method is used to make the TDLP to estimate the quarter hourly load profile of non-AMR customers. There were studies that examined the repeated clustering method in improving the ability to discriminate among the TDLP's of each cluster. Clustering techniques are exceptionally useful for assisting the distribution service providers in the process of classifying electricity customer on the basis of load pattern shape.

IV. EXISTING APPROACHES:

The Honey Bee Mating Optimization (HBMO) algorithm is a swarm-based type optimization technique, in which the searching procedure mimics the mating process in honey-bee colonies. Thus, the HBMO algorithm is related to the general field of swarm intelligence, but the mating process which is based on crossover and mutation operators, strongly relate this algorithm to evolutionary computing too.The mating-flight may be considered as a set of transitions in likely solutions where the queen moves among the similar states in some speed and mates with the drone encountered at each state probabilistically [12, 13, 14]. At the beginning of flight, each queen is initialized by an amount of energy and if this amount reaches a threshold or Zero, or even spermatheca has been filled, the queens will return to the nest. In this algorithm, workers' task is watching broods. In developed algorithm, workers are implemented as heuristic functions which cause fitness of broods to be increased.

A drone mates with a queen probabilistically using the function as *PQ-D*=exp(-|*FQ* – *FD*| / *SP*). where Prob(*Q*, *D*) is the probability of adding the sperm of drone *D* to the spermatheca of queen *Q* (that is, the probability of a successful mating); (*f*) is the absolute difference between the fitness of *D that is* *f*(*D*)) and the fitness of *Q that isf*(*Q*)); and *S*(*t*) is the speed of the queen at time*t*.It is apparent that this function acts as an-nealing function, where the probability of mating is high when either the queen is still in the start of her mating-flight or therefore her speed is high, or when the fitness of the drone is as good as the queen's. After each transition in space, the queen's speed *S* (*t*), and energy, *E* (*t*) decay using the following HBMO algorithm:

(i) The algorithm starts with the mating-flight, where a queen (best solution) selects drones probabilistically to form the spermatheca (list of drones).
(ii) A drone then selected from the list at random for the creation of broods. Creation of new broods (trial solutions) by cross- overing the drones' genotypes with the queen's.
(iii) Use of workers (heuristics) to conduct local search on broods (trial solutions).
(iv) Adaptation of workers fitness based on the amount of improvement achieved on broods.
(v) Replacement of weaker queens by fitter broods.

The algorithm starts with three user-defined parameters and one predefined parameter. The predefined parameter is the number of workers, representing the number of heuristics encoded in the program. The three user-defined parameters are the number of queens, the queen's spermatheca size and the number of broods that will be borne by all queens[15].

V. PROPOSED CASE-BO TO LOAD PROFILE CLUSTERING

STEP: 1 Unequal Random Cluster should be formed by using enhanced KNN Algorithm.
STEP: 2 Sustainability energy is considered for each tuple to analyze the lifespan of the cluster.
STEP:3 Analyze the Relationship announcement array of data.
STEP:4 Analyze the Relationship Delivery array of data .
STEP:5 Predict a Cluster Head for each unequal cluster by using the sustainability Value which is calculated in Phase 2.
STEP:6 Derive the distance value of each clusters .
STEP:7 Predict the Best Fit in every iteration using the above Steps.
STEP:8 End

The CASEBO algorithm is explained is as follows: Unequal Random Cluster should be formed by using enhanced KNN Algorithm.The sustainable energy **is** considered for each tuple to analyze the lifespan of the cluster.

$$\mathrm{ST(X)} = \frac{\mathrm{E_X,x}}{(P_{EH}, X+\alpha)Log}\left(\mathrm{E}^{\lambda(\mathrm{x})} - 1\right) \quad \text{--- 1}$$

$$\lambda(x) = \frac{\mathrm{E_x, x} - \mathrm{E}_y\mathrm{, x}}{\mathrm{E_x, x}}$$

where α *and* λ *are the constants* and Sustainable Energy is determined by ST(X), Where ST(X) is the individual energy capability of the node in cluster Association to sustain in Bees Optimization process.

Analyze the Relationship announcement array of data

$$ETX(k,d) = EelecK + Efskd^2 \quad \text{---2}$$

Analyze the Relationship Delivery array of data

$$ERX(k) = EelecK \quad \text{--- 3}$$

Predict a Cluster Head for each unequal cluster by using the sustainability Value which is calculated in Phase 2. Derive the distance value of each clusters by using this equation

$$\mathrm{D_C} = \sum_{j=1}^{\beta}\left[\left(\sum_{\mathrm{i=1}}^{\alpha_\mathrm{j}} \mathrm{d_{ijhj}}\right) + \mathrm{d_{hjca}}\right] + \sum_{j=1}^{\beta}\left[\left(\sum_{\mathrm{i=1}}^{\alpha_\mathrm{j}} \frac{\mathrm{ST\,(ij)}}{\mathrm{ST\,(hj)}}\right)\right]/\mathrm{k} \quad \text{--- 4}$$

Combine D_{ca} and E_c where E_c = E1+E2+E3.Where Dcais distance of cluster association, Ec isCollected Energy in Cluster Association CASE-BO = Dca+ E_c . Where the equation one(E1) is defined α *and* λ *are the constants* and Sustainable Energy is determined by ST(X), Where ST(X) is the individual energy capability of the node in cluster Association(CA) to sustain in Bees Optimization process. The equation two (E2) is defined the Transmission Energy with respect to random distance in the cluster Association . The equation three (E3) is defined the Receiver Energy with respect to random distance in the cluster Association . The equation four (E4) is**D**erive the distance value of each clusters and combine$D_{ca}$and E_c, that is E_c = E1+E2+E3 where collected energy in cluster association. Where Dca is distance of cluster association, E_c is Collected Energy in Cluster Association. CASE-BO = Dca+ E_c. . The above algorithm is proceeded to analyze the load profile of the electrical data and form an efficient load profile clustering algorithm.

The above equations use the following notations Where E_1 is the sum of Euclidean distances of cluster member to its cluster head and cluster heads to the Cluster Association(CA), Clp is a chromosome in the current round, αl (l = 1,…, β) is the number of cluster members, β is the number of clusters, dijhj is the Euclidean distance from node i in cluster j to its cluster head, dhjbs is the Euclidean distance from jth cluster head to the CA. Function E_2 is the ratio of the average energy sustainability of cluster members with its cluster head. Function E_3 is the ratio of the average Euclidean distance of the cluster heads to the CA with the sum of Euclidean distance of all the cluster nodes to the CA. So that the Cluster nodes are eliminated based on this threshold value which regains minimum iteration and Energy Efficiency. The fitness function defined above has the objective of simultaneously minimizing the intra-cluster distance between nodes and their cluster heads, as quantified by E_1 and of maximizing the cluster head's energy sustainability in its cluster as quantified by E_2; and of producing cluster with unequal size as quantified by E_3.

EXPERIMENTAL RESULT AND DISCUSSION:

The following snapshots are experimented randomly generated by the clustering preprocess for each categories of datasets.

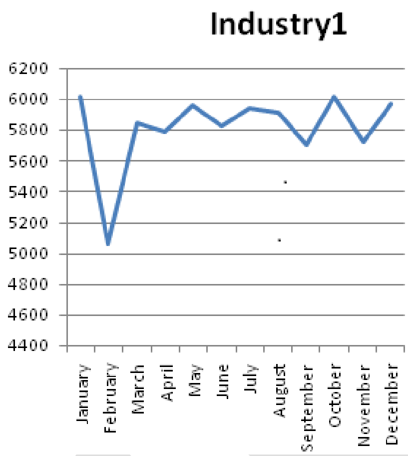

Figure 1

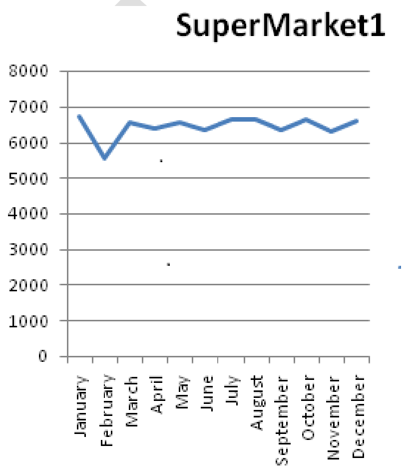

Figure 2

Figure-1 and Figure-2**d**escribes the Average data for each month after clustering and outlier data removal in Industry and Supermarket Sector

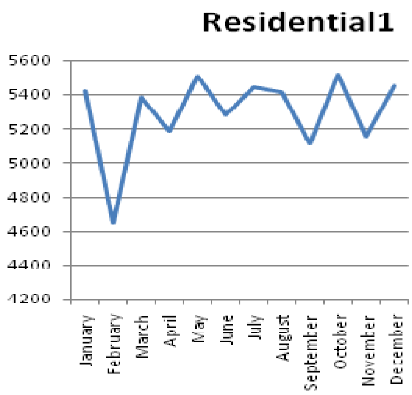

Figure-3

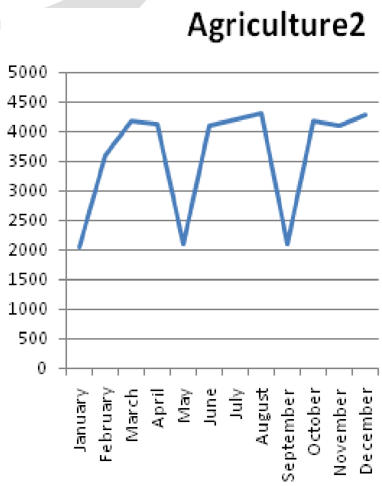

Figure-4

Figure-3and Figure-4describes the Average data for each month after clustering and outlier data removal in Residential and Agriculture Sector

Table1

	F	F1	F2	F3
Iteration1	**2064.919**	**10316.9**	**2.857143**	**0.04717**
Iteration2	**2371.678**	**11850.7**	**2.857143**	**0.049961**
Iteration3	**1836.071**	**9172.664**	**2.857143**	**0.050347**
Iteration4	**2923.844**	**14611.53**	**2.857143**	**0.053331**

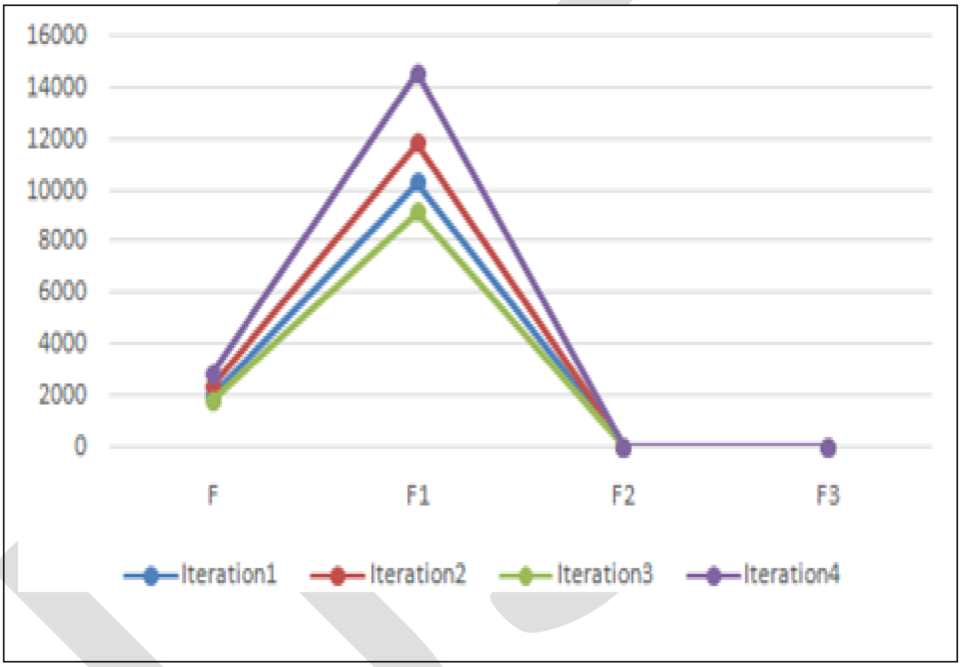

Figure-5 CASEBO

Figure-5 and Table 1 explains the 20 category of Large Scale Units with per day average Electricity Consumption Data with respect to Fitness F= (F1+F2+F3) and calculate the fitness value in each iterations of CASEBO.

Table-2

	F	F1	F2	F3
Iteration1	**1836.07137**	**9172.664**	**2.857143**	**0.050347**

Figure-6 and **Table 2** explains the 20 category of Large Scale Units with per day average Electricity Consumption Data with respect to Fitness F= (F1+F2+F3) and calculate the Best fitness value in each iterations of CASEBO.

Table 3

Category	Input Data(Average by Day)	CASEBO The Cluster Association Using Sustainable Energy Based		
		F1	F2	F3
Agriculture	136	8044.971	2.85714286	0.04659222
Bank	174.5			
Hospital	192			
Hostel	189			
Hotel	211			
Industries	450	F=1610.53202986189		
Residential	149.5			
Supermarket	273			
Theater	471.5			
University	170.5			

Table 3 explains average fitness for all categories for per day.

The above data's were collected for a Year and used to run the CASEBO algorithm in different fitness values as per the Indian context. Based on the representation of attributes the TLPs produced by the CASEBO algorithm for the consumer categories. We conclude that the optimal solution for the LP clustering problem discussed in this paper and uses a chromosomes encoding scheme with 20 clusters/TLPs and a fitness function computed. Results are comparable to others, the new LP clustering method based on CASEBO algorithm can be easily implemented with high robustness properties. The dataset categories are I)Residential two substations (Resi-1 and Resi-2), ii) Industrial (Ind-1 and Ind-2), iii) Agriculture (Agri-1 and Agri-2) iv) Commercial such as seven categories are Hospital, super market, Theatre, University, Bank, Hostel, Hotel (each two categories). The dataset were specially considered to the use of fitness function with twenty categories of services. The overall dataset is evaluated by the attributes, the attribute statistics representations are as follows. DATE, FROMTIME, TOTIME, TYPES, SERVICEID, KWATTS.

VI. CONCLUSION:

This paper presents an efficient method for the classification and load profiling of distribution network customers. The proposed method was implemented as a MATLAB program and tested with real data and also it was tested with weka. The result showed that theCASEBO algorithm can classify customers into well separated clusters according to their electricity consumption data, and clearly indicate that the proposed CASEBO algorithm has significant implications with its efficient and stable nature of the structure in handling the database. Therefore, it is evident that new LP clustering approach has the potential in efficiently addressing metering issues. Correspondingly, with little efforts by manipulating the required parameters it is possible to obtain the desired results and reach the goal. In light of this result, it is certainly possible to extend further on these investigations to develop Enhanced Replacement inBest Optimization Algorithm where in both TLPs and the number of clusters with simultaneous fitness function integrated in the investigation of metering.

It was proved earlier that, the resulting customer classification is more accurate than the alternative classification methods. The snapshots are clearly indicated that the proposed CASEBO Clusters and load profile data more accurately than the existing methods which was also implemented and it also proves that CASEBO is also accurate than the most existing approaches.

VII. REFERENCES:

[1]. Application of Honey Bee Mating Optimization algorithm to Load Profile Clustering. MihaiGavrilas ”Gh.Asachi” Technical University of Iasi, Romania,978-1-4244-7230-2010-IEEE.

[2].V.Figueiredo,F.Rodrigues,Vale, and J.B.Gouveia,”An electricenergy consumer characterizationframework based on data mining techniques”, IEEE trans.power syst., vol.20, No.2, 2005, pp.596-602.

[3]. E. Hirst and B. Kirby, Retail Load Participation in Competitive Wholesale Electricity Markets. Washington, DC: Edison Electric Institute Rep., 2001.

[4]. B. Pitt and D. Kirchen, “Applications of data mining techniques to load profiling,” in Proc. IEEE PICA, Santa Clara, CA, May 1999, pp. 131–136.

[5] .C. S. Chen, J. C. Hwang, and C.W. Huang, "Application of load survey to proper tariff design," IEEE Trans. Power Syst., vol. 12, no. 4, pp. 1746–1751, Nov. 1997.

[6]. G. Chicco, R. Napoli, P. Postulache, M. Scutariu, and C. Toader, "Customer characterization options for improving the tariff offer," IEEE Trans. Power Syst., vol. 18, no. 1, pp. 381–387, Feb. 2003.

[7] .V. Figueiredo, F. J. Duarte, F. Rodrigues, Z. Vale, and J. Gouveiaet al., "Electric energy customer characterization by clustering," in Proc. ISAP, Lemnos, Greece, Sep. 2003.

[8] .U. Fayyad, G. Piatetsky-Shapiro, P. J. Smith, and R. Uthurasamy, "From data mining to knowledge discovery: an overview," in Advances in Knowledge Discovery and Data Mining. Cambridge, MA: AAAI/MIT Press, 1996, pp. 1–34.

[9]. F. Rodrigues, V. Figueiredo, F. J. Duarte, and Z. Vale, "A comparativeanalysis of clustering algorithms applied to load profiling," in LectureNotes in Artificial Intelligence (LNAI 2734). New York: Springer-Verlag, 2003, pp. 73–85.

[10].G.Chicco, R.Napoli,andF.Piglione, "comparisons among clustering techniques for electricity customer classification," IEEE trans.power syst., vol.21,No.2,pp.933-940,May 2006.

[11].M.Gavrilas, O.Ivanov, G.Gavrilas,"Load profiling with fuzzy self-organizing", proceedings of the 9th symposium on neural network applications in electrical engineering NEUREL 2008,Belgrade,sept.25-27,2008,CD-ROM,ISBN:978-1-4244-2903-S.

[12].S.V.Verda,M.O.Garcia,C.Senabre,A.G.Marin,F.J.G.Franco:Classification, filtering and identification of electrical customer load patterns through the use of self-organizing maps,"IEEE trans power syst.,vol.21,no.4,pp.672-1682,Nov.2006.

[13].D.Gerbec, S.GAsperic, I.Smon,andF.Gubina, "Determining the load profiles of consumers based on fuzzy and probability neural networks," proc. Inst. Elect Enmg., Gener., Trans., Distrib., vol. 151,no.3,pp.395-400, May.2004.

[14].Hirst and B.Kirby,Retail Load Participation in competitive Wholesale Electricity Markets. Washington, DC:Edison Electric Institute Rep,2001

[15]. B.pitt and D.kirchen ,"Applications of data mining techniques to load profiling ," in proc .IEEE PICA, santaClara ,CA,May 1999, pp.131-136