A NEW APPROACH FOR CALCULATING LOAD AND LOSS FACTOR BASE ON CONSUMER DATA WITH FUZZY MODELLING.

Shahram KAYKAHIE
Iran
it@kedc.ir

Sina KOWSARI MOVAHED
Iran
ict@kedc.ir

ABSTRACT

In Iran nowadays, electric distribution utilities observe the considerable increase of the technical and non-technical losses in their network. Thus utilities need to look for more accurate tools for loss calculation. In order to select the optimal solution for loss reduction, need to have loss estimation or calculation models. Given that the total amount of losses in a distribution system is known, with a reliable methodology for the technical loss calculation, the non-technical losses can be obtained by subtraction. One way to calculate the loss value is calculating loss factor base on load factor. Since computing load factor needs metering the peak of electric power Demand this metering in large scale distribution network have more charge and is expensive. In this paper we introduce the methods that computing these factors, without metering the peak of electric power Demand, based on fuzzy clustering of monthly energy consumption of consumers in billing database.

INTRODUCTION

We are going to present a new technique that based on historical energy consumption data of consumer, Load and Loss factor calculated. In order to use this method the different customers must be classified and coincident peak demand and annual energy consumption for each class must be available. Since measuring annual coincident peak demand not easy, we are using estimated values. The estimation of annual coincident peak demand has traditionally been performed by different empirical methods. Research shows that if the customer loads in peak time are normally distributed the total peak load can be computed by a method using statistical analysis and typical peak load for each class of customer [1-3]. We concentrate on the discovery of association rules by clustering consumption data base on fuzzy c-means by particle swarm optimization and then generating normal rules. The first and most intuitive approach implements all possible combinations of the given fuzzy sets as rules. There are numerous methods to define the membership functions. We considered consumption of twelve month of each customer in selected zone and then normal rules calculated in data that studied. This paper aims to introduce the system that base on profile of consumer per kWh normal rules generated and for each rule load factor and loss factor calculated. So first we introduce clustering method and fuzzy c-means clustering by particle swarm optimization next in second section we introduce fuzzy membership function generation. In the third section load modeling base on monthly energy consumption introduce. And in final section we introduce heuristic algorithm and two case study for calculating load and loss factor.

CLUSTERING METHODS

Cluster analysis is a term used to describe a family of statistical procedures specifically designed to discover classifications within complex data sets. The objective of cluster analysis is to group 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. Thus, the purpose of the analysis is to arrange objects into relatively homogeneous groups based on multivariate observations.

Fuzzy c-means algorithm

Given a number of historical examples of a monthly measured kWh $[x_1, x_2, \dots, x_N]$ and the anticipated number of clusters, c, the Fuzzy c-means (FCM) algorithm partitions these examples into c clusters by minimizing the objective function Q:

$$Q = \sum_{k=1}^{c} \sum_{i=1}^{N} u_{i,k}^{m} \| x_i, o_k \|^2 \qquad (1)$$

In this objective function, Q represents the sum of the distance of individual data to the cluster centers $[o_1, o_2, \dots, o_C]$ the cluster centers are also called prototypes). Here $u_{i,k}$ represents the membership of data x_i belonging to cluster k. The parameter m stands for the fuzzification factor that takes the value of 2 in most of applications of FCM. The FCM iteratively updates the prototypes $[o_1, o_2, \dots, o_C]$ and memberships $u_{i,k}$ through the equations below [6]:

$$o_k = \frac{\sum_{i=1}^{N} u_{i,k}^{m} . x_i}{\sum_{i=1}^{N} u_{i,k}^{m}} \qquad (2) \qquad\qquad u_{i,k} = \frac{1}{\sum_{j=1}^{c} \left(\frac{x_i - o_k}{x_i - o_j} \right)^{\frac{2}{m-1}}} \qquad (3)$$

Iteration will stop when:

$$\max(| u_{i,k}^{new} - u_{i,k}^{old} |) < e \qquad (4)$$

Where *e* is a termination criterion. This procedure helps to find values of prototypes and memberships that achieve a saddle point or a local minimum of the objective function *Q*.

Particle Swarm Optimization

Fuzzy c-means is supervisor learning algorithm and in this paper for selecting optimal count of each cluster we use Particle Swarm Optimization. PSO [4-5] optimizes an objective function (Eq-11) by undertaking a population – based search. The population consists of potential solutions, named particles, which are metaphor of birds in flocks. The various steps involved in Particle Swarm Optimization Algorithm are as follows [5].Where X and V are position and velocity of particle respectively. w is inertia weight, c1 and c2 are positive constants, called acceleration coefficients which control the influence of p_{best} and g_{best} on the search process, P is the number of particles in the swarm,r1 and r2 are random values in range [1 , 0] .

1. Initialize the parameters including population size P, c_1, c_2 w, and the maximum iterative count.
2. Create a swarm with P particles (X, p_{best}, g_{best} and V are n×c matrices).
3. Initialize X, V, p_{best} for each particle and g_{best} for the swarm.
4. Calculate the cluster centers for each particle using Eq. (2).
5. Calculate the fitness value of each particle using Eq. (11).
6. Calculate p_{best} for each particle.
7. Calculate g_{best} for the swarm.

$$p_{i,best} = p_i \quad if\ f(p_i) > f(p_{i,best}) \tag{5}$$

$$g_{i,best} = g_i \quad if\ f(g_i) > f(g_{i,best}) \tag{6}$$

8. Update the velocity matrix for each particle using

$$v_i = wv_i + c_1R_1(p_{i,best} - p_i) + c_2R_2(g_{i,best} - p_i) \tag{7}$$

9. Update the position matrix for each particle using:

$$p_i = p_i + v_i \tag{8}$$

10. If terminating condition is not met, go to step 4.

PSO utilizes several searching points and the searching points gradually get close to the global optimal point using its p_{best} and g_{best}. Initial positions of p_{best} and g_{best} are different. However, using the different direction of p_{best} and g_{best}, all agents gradually get close to the global optimum.

The Fuzzy Cluster Validity Criterion

As a commonly used fuzzy cluster validity function, the partition coefficient measures the amount of overlap among clusters. But the disadvantages of partition coefficient are the lack of direct connection to geometrical property and its monotonic decreasing tendency with c [7].The global compactness of the fuzzy c-means of the input data set can be calculated from:

$$\sigma = \frac{1}{n}\sum_{i,k} \mu_{i,k}^{m} ||x_k - o_i||^2 \tag{9}$$

Therefore, the better c-means, the smaller σ. The separation among clusters can be calculated by:

$$\pi = \min_{i,j} ||o_i - o_j|| \tag{10}$$

It is the separation of fuzzy c-means. Obviously, if all the fuzzy clusters are completely separated, leading to the largest π.We can then define the fuzzy cluster validity to be[7]:

$$S = \frac{\sigma}{\pi} = \frac{\frac{1}{n}\sum_{i=1}^{c}\sum_{i,k}\mu_{i,k}^{m}||x_k - o_i||^2}{\min_{i,j}||o_i - o_j||} \tag{11}$$

As a result, the separation among clusters and compactness within clusters is measured to give a fuzzy validity criterion based on a validity function witch identifies overall compactness and separate fuzzy c-portion without assumption as to the number of substructures inherit in the data.

MEMBERSHIP FUNCTION GENERATION BASED ON FCM

After we obtain the cluster centers and membership matrix ($u_{i,k}$), there are numerous methods to define the membership functions based on it. To define membership functions using FCM's results discards all the membership values and just uses the centers of the clusters $[o_1, o_2, \dots, o_C]$ to define specific shaped membership functions, such as Gaussian and triangular functions. Fig1 demonstrated this method. The trapezoidal and triangular membership functions are defined by prototypes generated in FCM algorithms.

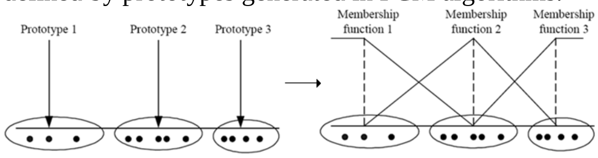
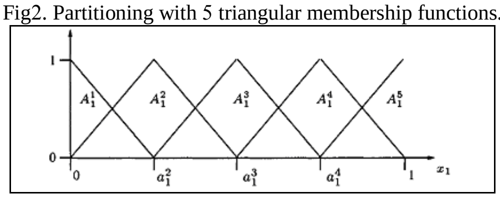


Fig1 FCM-Centers/Prototype-base Membership Function

CLASSIFICATION OF CONSUMER

We assume the input and output spaces to be $[0, 1]^P$ and $[0, 1]$. For the iTh input variable x_i, its domain interval is evenly divided into K_i fuzzy sets labelled as $A_i^1, A_i^2, A_i^3, \dots A_i^{K_i}$, as shown in Fig.2 for the variable with K_1= 5. Membership function can be used, the most common being the triangle-shaped one All the rules corresponding to the possible combinations of the inputs are implemented. The total number of rules for a input system is: $K_1 \times K_2 \times K_3 \times \dots \times K_p$.[4]

Fig2. Partitioning with 5 triangular membership functions.



On the other hand in each input (monthly consumptions profile) of all users clustered and then, finding association rules in databases with fuzzy attribute. We introduce the fuzzy association rules of the form, 'If X is A and Y is B then consumer is normal', to deal with quantitative attribute, and if this rule have support greater than x must tagged as normal rule [4].

Table 1.Sample of normal rules:

$if\ A_1^1 \& A_2^1 \& A_3^1 \& A_4^1 \& A_5^2 \& A_6^1\ then\ normal$

$if\ A_1^2 \& A_2^1 \& A_3^2 \& A_4^2 \& A_5^3 \& A_6^2\ then\ normal$

The idea of modelling is to create load models that describe the electricity consumer's consumption of similar class of consumers. For this purpose, all customers are divided on several groups, in which the monthly electricity consumption profile can be assumed to be equal with a sufficient accuracy. Calculating coincident peak demand, during that it was consumed have to be done for each customer group.

LOAD MODELING BASED ON THE ANNUAL ENERGY CONSUMED DATA

If real-time measurements of power and current in the end-user's side are not available in the distribution network, a value of the customer's annual energy is taken as a staring for the customer load modelling. The amount of annual energy on each customer is taken from the electricity billing database then the value of customer annual energy has to be converted into the peak power (VALENDER's formula) or into the power at a specified instant (load models). The methods of conversions are typically based on information about the customer load behaviour. VALENDER's formula (Eq-12) was a conventional method to estimate peak loads in distribution networks[2-3].

$$P_{max} = k_1 . W + k_2 \sqrt{W} \quad (12)$$

$$P_{max} = peak\ power, kW \quad (13)$$

$$W = annual\ energy\ of\ the\ customer, MWh \quad (14)$$

$$k_1 . k_2 = Velander\ coefficients \quad (15)$$

$$P_{max} = k_1 . W + k_2 \sqrt{W} \quad (16)$$

The factors k_1 and k_2 are empirical coefficients and examples are given in table 2. Also possible maximum demands per customer depending on the annual energy consumption are included. VALENDER's formula are based on stochastic formulas and can be expressed in the equations:

$$P_{mean,n} = n . P_{mean,1} \quad (17)$$

$$\sigma_n = \sqrt{n} . \sigma_1 \quad (18)$$

The maximum loads do not occur at the same moment. For specific load groups it can be assumed that the various loads during a peak period are normally distributed around a certain mean value. Each load for a certain moment can then be represented with a mean value and a standard deviation. With c being a factor describing the relation between the extreme value, the mean value and the standard deviation:

$$c = \frac{P_{max,1} - P_{mean,1}}{\sigma} \quad (19)$$

$$P_{max,n} = P_{mean,n} + c . \sigma_n = n . P_{mean,1} + (P_{max,1} - P_{mean,1}) . \sqrt{n} \quad (20)$$

The coincidence factor g_n is then:

$$g_n = \frac{P_{max,n}}{n . P_{max,1}} = \frac{P_{mean,1}}{P_{max,1}} + (1 - \frac{P_{mean,1}}{P_{max,1}}) \frac{1}{\sqrt{n}} \quad (21)$$

$$g_\infty = \frac{P_{mean,1}}{P_{max,1}} \quad (22)$$

$$g_n = \left(g_\infty + (1 - g_\infty) \frac{1}{\sqrt{n}} \right) \quad (23)$$

The coincident peak demand is calculated with $g_\infty = 0.2$.

$$P_{max,n} = \alpha . n . W + \beta \sqrt{n . W} \quad (24)$$

$$\alpha = \frac{P_{max,1} . g_\infty}{W} \quad (25)$$

$$\beta = \frac{P_{max,1} . (1 - g_\infty)}{\sqrt{W}} \quad (26)$$

Table 2. Empirical coefficients and examples

class	k_1	k_2	E (kwh)	$P_{max,1}(kw$	$P_{max}(100)$
Domestic	0.00033	0.05	2000	2.9	81
			3500	4.1	115
			5500	5.2	145

LOAD AND LOSS FACTOR DEFINITIONS

One of the important data for the correct measurement of losses in distribution networks are the load factor and the loss factor, which present a strong influence on the determination of loss unitary costs. The load factor is usually obtained with energy and demand measurements, whereas, to compute the loss factor it is necessary the learning of demand and energy loss, which are not, in general, prone of direct measurements [1-2].

Load Factor (LF):

Ratio between the average power ($D_{average}$) and the maximum demand (D_{max}), in monthly measurement where:

$$LF = \frac{D_{average}}{D_{max}} = \frac{1}{D_{max}} * \frac{\sum_{i=1}^{T} D(i)}{T} \quad (27)$$

$$LF = \frac{E}{T * D_{max}} \quad (28)$$

$\sum_{i=1}^{T} D(i)$ is the monthly demand that represents the energy supplied to the system (E) during the period of the time T. Thus, (27) is obtained. In this paper we consider coincident peak demand energy at each normal rule and calculating from (24). In the other hand:

$$LF = \frac{E}{T * D_{max}} \cong \frac{E}{T * (\alpha . n . W + \beta \sqrt{n . W})} \quad (29)$$

Loss Factor (LLF):

Ratio between the average power losses ($L_{average}$) and the losses during peak load (L_{max}), in a period of the time with monthly measurement. In other words, the loss factor is simply the load factor of the losses.

$$LLF = \frac{L_{average}}{L_{max}} = \frac{e}{T * L_{max}} \quad (30)$$

Where $\sum_{t=1}^{T} L(i)$ represent the energy losses of the system (e) during the period of the time T. Since the late '20s, researchers have been looking for a form to relate the loss factor (LLF) with the load factor (LF). All the studies led to an empiric equation [1].

$$LLF = (LF) * 0.3 + (LF)^2 * 0.7 \quad (31)$$

However the load and loss factors could be obtained in a better manner by distribution utilities if it is used the load duration curve of each class of consumer. The way used in our algorithm is the estimation of peak load from Eq.24. By considering monthly energy profile and peak demand of each class of consumer we are perform load flow and calculating peak loss L_{max} in low voltage network that studied.

$$e = LLF * L_{max} \quad (32)$$

ALGORITHM

1. Select measured consumption per kWh of any consumer with 2 month length in six duration in year.
2. Normalized data and map consumption data to sixty day.
3. Clustering each duration data with fuzzy c-means.
4. Automatically generate membership functions and then calculate Cartesian relation as normal rules.
5. Calculated count of consumer that complied normal rules.
6. Calculated LF base on average of consumption and greatest member function for each normal rules.

7. Calculated average of LF of all profile.
8. Calculated LLF in each duration base on LF.
9. Perform load flow and calculated peak loss in low voltage network and by Eq. (32) we have calculated average power losses in our network.

Case study1. Consider we need to calculate LF and LLF in a LV-Feeder with 125 consumers with the same consumption that is together in same cluster. The result showing in table 3.

Table 3. Calculating LF and LLF base on generated rules

First Period (Kwh)	**Second Period (kwh)**	**Sum (kwh)**	$avg(mf_i)_i$	**Population**
mf_1^1=250	mf_1^2=350	$sum_1 = ٦٠٠$	avg_1 =300	C_1=20
mf_2^1=250	mf_2^2=700	$sum_2 = ٩٥٠$	avg_2 =475	C_2=25
mf_3^1=750	mf_3^2=350	$sum_3 = ١١٠٠$	avg_3 =550	C_3=30
mf_4^1 =750	mf_4^2=700	$sum_4 = ١٤٥٠$	avg_4 =725	C_4=50
Calculating Coincidence peak $P_{max} = k_1 . sum_i + k_2 \sqrt{sum_i}$				**kw**
$P_{max1} = ٢٠ * ٠.٠٠٠٣٣ * ٦٠٠ + ٠.٠٥ * \sqrt{٢٠ * ٦٠٠}$				5.9
$P_{max2} = ٢٥ * ٠.٠٠٠٣٣ * ٩٥٠ + ٠.٠٥ * \sqrt{٢٥ * ٩٥٠}$				9.4
$P_{max3} = ٣٠ * ٠.٠٠٠٣٣ * ١١٠٠ + ٠.٠.٥ * \sqrt{٣٠ * ١١٠٠}$				11.9
$P_{max4} = ٥٠ * ٠.٠٠٠٣٣ * ٧٥٠ + ٠.٠٥ * \sqrt{٥٠ * ١٤٥٠}$				21.5
$LF_{avg} = ٠.٥٦$				
$\dfrac{\frac{avg_1 * ٢٠}{P_{max1} * ٢٤ * ٦٠} + \frac{avg_2 * ٢٥}{P_{max2} * ٢٤ * ٦٠} + \frac{avg_3 * ٣٠}{P_{max3} * ٢٤ * ٦٠} + \frac{avg_4 * ٥٠}{P_{max4} * ٢٤ * ٦٠}}{٤}$				
LLF calculating				
LLF=$LF_{avg} * ٠.٣ + LF_{avg}^{٢} * ٠.٧ = ٠.٣٨$				

Case study2. There are 512 residential customer connected to distribution transformer with the monthly consumed energy profile that shows in table 3. Assume that connected load is 5.5kW per house .Determine Load and Loss factor for this group of consumer that connected to low voltage feeder of specified distribution transformer. Table 3 shows sample of generated rules and abstract of calculation for load and loss factor. In this table load factor 20 % for total of consumer calculated and then loss factor equal 9 % calculated.

CONCLUSION

Nowadays in electricity distribution companies computing load factor needs metering the peak of electric power Demand .This metering in large scale distribution network have more charge and is expensive. Therefore we need heuristic method for estimation peak factor. Since there is not proper algorithm for modeling loss in electrical network, in this paper by using fuzzy clustering by particle swarm optimization we introduce a system for classification of consumer in distinct class based on energy consumption profile. In this method we use VALENDER's formula (Eq.12) for estimation coincident peak demand of any class of consumer. For example the result in table 3 shows that four class of consumer with different population have total load factor 0.56 and total loss factor 0.38. In second case study we studied specific region with 554 consumers that connected to special low voltage feeder. We have calculated sixteen class of consumer with different class and population. By using eq.12 we have calculated coincident peak demand and load factor for each class of consumer then by calculating weighted average of each load factor we have calculated whole load factor 21% for all consumer in low voltage feeder. And in final we have calculated total loss factor 9%.

REFERENCES

[1].M. E. de Oliveira, A. Padilha-Feltrin, and F. J. Canadian "Investigation of the Relationship between Load and Loss Factors for a Brazilian Electric Utility" 2006 IEEE PES Transmission and Distribution Conference and Exposition Latin America, Venezuela.
[2]. F. Provoost, "Intelligent distribution network design," 2009, p. 333.
[3]. J. Dickert, P. Schegner" Residential Load Models for Network Planning Purposes" Modern Electric Power Systems 2010, Wroclaw, Poland
[4] .R.C.Eberhart and Y.Shi (2001), "Particle Swarm Optimization: developments, applications and resources," Proceedings Congress on Evolutionary Computation, IEEE Service Centre, Piscataway, NJ, Seoul, Korea.
[5].Shi.Y and Eberhart.R.C (1998), "A modified particle swarm optimizer", Proc. IEEE International Conference on Evolutionary Computation, IEEE Press, Piscataway, NJ, pp. 69-73.
[6] F. Hoppner, F. Klawonn, R. Kruse, and T. Runkler, Fuzzy ClusterAnalysis. Chichester, Wiley, 1999.
[7] X.Xionb,K.L.Chan, "Towards An unsupervised Optimal Fuzzy Clustering Algorithm for Image Database Organizaion:,2000,IEEE.

Table 3. Calculating LF and LLF base on sample generated rules on second study

Cluster Centers														
	68	95				C1								
454	545	395	705	248	57	C2								
125	293	777	205	468	202	C3				Eq(24-26)				
	993	1476		69	373	C4	average		Eq.12($k_1 = 0.00033$, $k_2 = 0.05$)	$g_{-} = ٠.٢$			Eq.29	Eq.31
Period														
1	2	3	4	5	6	count	W(kwh)	W(kwh)		Alpha	Beta	Pmax,n	LF	LLF(n)
125	68	95	205	69	57	129	103.2	619	1.45	0.00047	0.047	50.7	0.18	23.22
125	293	395	205	248	202	67	244.7	1468	2.4	0.00033	0.05	47.8	0.23	15.41
454	545	777	705	468	373	60	553.7	3322	3.98	0.00024	0.055	72.4	0.31	18.6
454	545	777	705	248	202	37	488.5	2931	3.67	0.00025	0.054	44.9	0.28	10.36
454	293	395	205	248	202	37	299.5	1797	2.71	0.0003	0.051	33.2	0.23	8.51