A NEW METHOD FOR OPTIMAL PLANNING IN EXTENSIVE DISTRIBUTION NETWORKS DESPITE UNCERTAIN DATA

Mostafa FALAHI, Daneshmand co – Iran, ms.falahi@gmail.com
Hosein ALAVI, aepdco – Iran, h.alavi@yahoo.com
Mastoore SHAYGAN, Kzrec – Iran, ms.shaygan@gmail.com
Ashcan HOJATI, Daneshmand co – Iran, edsu_daneshmandco@yahoo.com

ABSTRACT

The distribution network planning is a multi-objective problem including minimizing losses and investment. By setting planning priorities, the necessary actions to loss reduction are prioritized, thus the waste of investment will be prevented. In this paper, a new method is presented to determine planning priorities of the extensive and complex distribution network with uncertain data to optimize network variable including power losses and voltage profile.

INTRODUCTION

Optimal utilization of the distribution network and providing sufficient capacity to meet load growth, entail optimal planning of distribution network.
In distribution network planning to improves network variable the necessary actions to modifying network parameters including *connected load*, *transformer installed capacity and length of distribution line*, are prioritized .
Some of these actions are the following [2]:
- Identification of the weakest areas in distribution network and improving them;
- Reduction of the length of the distribution networks by relocation of distribution substation/installations of additional transformers;
- Identification overloads feeders, and reallocates the loads.

In an extensive distribution network, with uncertainty in the data it is impossible to carry out an exhaustive analysis [1], thus, the use of statistical methods that are designed to discover the complex relationships between the available data would be so effective. One of these methods is Cluster analysis that as an unsupervised process divides a set of objects into homogeneous groups.
Data clustering (or just clustering), also called cluster analysis, segmentation analysis, taxonomy analysis, or unsupervised classification, is a method of creating groups of objects, or clusters, in such a way that objects in one cluster are very similar and objects in different clusters are quite distinct. Generally, the common sense of a cluster will combine various plausible criteria and require [5], for example, all objects in a cluster to
- Share the same or closely related properties;
- Show small mutual distances or dissimilarities;
- Have “contacts” or “relations” with at least one other object in the group; or
- Be clearly distinguishable from the complement, i.e., the rest of the objects in the data set.

The conventional k-means algorithm is one of the most used clustering algorithms that have some important properties:
- It is efficient in clustering large data sets, since its computational complexity is linearly proportional to the size of the data sets.
- It often terminates at a local optimum.
- The clusters have convex shapes, such as a ball in three-dimensional space.
- It works on numerical data.
- The performance is dependent on the initialization of the centres

According to the properties of k-means algorithm, here we use this cluster analysis to indicate the parameters affecting distribution network variables *including power losses and voltage profile*, and the results will be used to determine the scheduling priorities.

K-MEANS CLUSTERING ALGORITHM:

The K-means clustering is an algorithm to classify set of data based on attributes into K number of group. The main idea behind the global k-means algorithm is that an optimal solution of a clustering problem with k clusters can be obtained by carrying out a series of local searches using the k-means algorithm. The k-means algorithm can be divided into two phases: the initialization phase and the iteration phase. In the initialization phase, the algorithm randomly assigns the cases into k clusters. In the iteration phase, the algorithm computes the distance between each case and each cluster and assigns the case to the nearest cluster.
We can treat the k-means algorithm as an optimization problem. In this sense, the goal of the algorithm is to minimize a given objective function under certain conditions.
The global k-means algorithm is described as follows:
Let D ={x_1, x_2,..., x_n} be a given data set in a d-dimensional space. The aim of the clustering problem with k clusters is to partition D into k clusters C_1, C2, ... , Ck , such that the clustering criterion

$$E(\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_k) = \sum_{j=1}^{k} \sum_{\mathbf{x} \in C_j} d_{euc}^2(\mathbf{x}, \mathbf{m}_j)$$

is optimized, where m_j is the centroid of cluster C_j and d_{euc}(.,.) is the Euclidean distance[5].
The algorithm begins by initializing a set of K cluster centres. Then, it assigns each object of the dataset to the cluster whose centre is the nearest, and recomputed the centres. The process continues until the centres of the clusters stop changing.
In order to speed up the global k-means algorithm, some techniques, such as initialization with kd-trees, are used. Hussein [8] proposed another global k-means algorithm based on a greedy approach, called the greedy k-means algorithm, in order to avoid some drawbacks of the global k-means algorithm.
It is obvious in this algorithm that the final clusters will

depend on the initial cluster centres chooses and on the values of K.
For determination the optimal number of clusters; properties of clusters such as density, sizes and form of cluster, separability of clusters, robustness of classification, will be evaluated using *Silhouette Global Index* [6]. Using this approach each cluster could be represented by so called silhouette, which is based on the comparison of its tightness and separation. The average silhouette width will be applied for evaluation of clustering validity and also will be used to decide determination of optimal number of clusters.

$$SC = \frac{1}{N_c} \sum_{j=1}^{N_c} S_j$$

Where:
S_j silhouette local coefficient is defined as:

$$S_j = \frac{1}{r_j} \sum_{i=1}^{r_j} s_i$$

s_i the silhouette width index for i-object is:

$$s_i = \frac{b_i - a_i}{\max\{b_i \ , a_i\}}$$

a_i – mean distance between object i and objects of the same class j,
b_i – minimum mean distance between object i and objects in class closest to class j.
In [7] is proposed the following interpretation of the SC coefficient:
• 0.71 – 1.0 A strong structure has been found;
• 0.51 – 0.7 A reasonable structure has been found;
• 0.26 – 0.5 the structure is weak and could be artificial;
• < 0.25 No substantial structure has been found.
Cluster validity checking is one of the most important issues in cluster analysis related to the inherent features of the data set under concern. Validity indices are measures that are used to evaluate and assess the results of a clustering algorithm; it aims at the evaluation of clustering results and the selection of the scheme that best fits the underlying data.

OPTIMAL DISTRIBUTION NETWORK PLANNING:

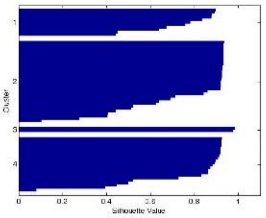
The metering systems provide the necessary information for network planning, but in practice, we often encounter two problems in the metering systems: some important data are missing in the data sets, and there might be errors in the data sets.
Methods that be used in network planning, should be able to deal with missing values and errors in data, and provide appropriate decisions in conditions with uncertainty data. Thus the waste of investment as power loss and plans that do not help to solve the network problems will be prevented.
According to the Properties of k-mean cluster algorithm, this method can be used to optimal distribution network planning with uncertainty data.
In the proposed method, measurements will be collected as input of network planning, and using k-mean cluster algorithm relationship between them and network parameters will be discovered, then this relationship will be prioritized based on the severity of dependence, thus parameters that have the greatest impact on improving the network variables are identified and use as output of network planning.

CASE STUDY:

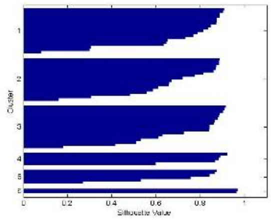
To illustrate the effectiveness of the proposed method the network of Lashkar Abad in Ahwaz power distribution companies has been studied.

Fig 1: Lashkar Abad distribution network, there are three 11KV feeder that shown with different color, there are also 69 transformer that depicted with circle around them

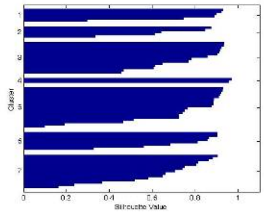
This network is a wide area network with 69 transformers and an 18.8 km LV feeder.
Firstly due to feeding area of the distribution transformers, the geographical network area is divided into separate blocks.
Then, Variables and network parameters are calculated for each of the blocks, Network parameters in various blocks are different and Power losses and voltage profiles depend on these parameters.
Then, Blocks with the similar dependency between losses and voltage profile to network parameters, are classified in the same group By K means clustering algorithms, the global silhouette coefficient is calculated to determine the optimal number of cluster (K).
Since the dependency among various parameters and network variable are not the same, Cluster analysis is performed separately for each of the network parameters, And the optimal number of clusters for each parameter is determined according to silhouette Index. By changing the number of clusters, the silhouette index will also change. In the optimal case, the index is higher than 0.6, for the majority of members of each clusters.
As Shown in Figure 2, the optimal number of clusters, to describe the relationship between lengths of distribution line, connected load and transformer installed capacity, with the power losses, are respectively 4, 7 and 6 and to describe the relationship between the length of distribution line and the voltage drops the optimal number of cluster is 7.



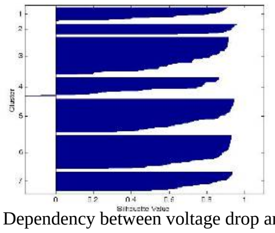
Dependency between power losses and LV feeder length (k=4)



Dependency between power losses and trans installed cap (k=6)



Dependency between power losses and connected load (k=7)



Dependency between voltage drop and LV feeder length (k=7)

Fig 2: The Silhouette Index plot for various parameters

After determining optimal number of clusters for various parameters, the blocks will be classified using the K-means algorithm.
Classified blocks are shown in the following figures, central parameters and variables are written beside each category.

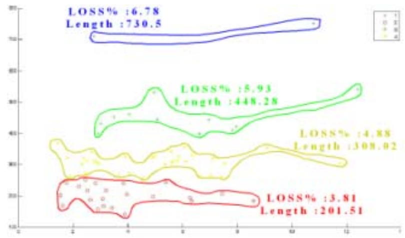


Fig 3: Clustered blocks based on dependency between power losses and LV feeder length

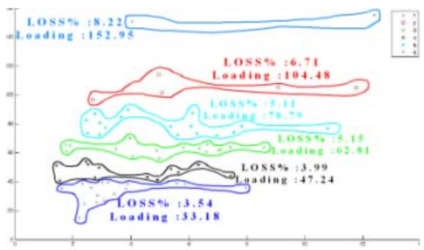


Fig 4: Clustered blocks based on dependency between power losses and trans installed cap

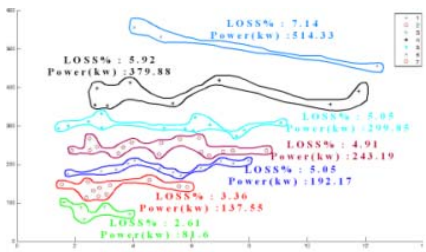


Fig 5: Clustered blocks based on dependency between power losses and connected load

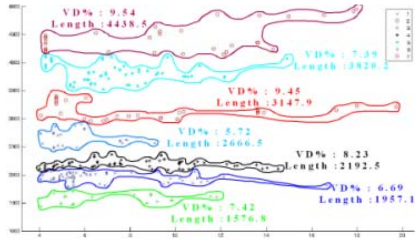


Fig 6: Clustered blocks based on dependency between voltage drop and LV feeder length

In order to consider the uncertainty of the data, ranges for variables and network parameters are defined, and non-deterministic expressions are used to describe each range; table 1 shows an example of non-deterministic expressions for power losses. Other variables and parameters will have their own ranges. (The phrase “Very high" describing the critical range)

Table 1: non-deterministic expressions for power losses

expression	Variable range (%)
Low	power loss < 4.5
High	4.5 < power loss < 6
Very high	6 < power loss

Identify critical blocks:

According to the results of clustering algorithm, the cluster that its central variable and parameter is located in the range of the phrase "very high", specifies the blocks with highly interdependent variables and parameters, this show that the variables of this blocks (losses and voltage profiles) are in the critical range and take actions to correct the specified parameters in this blocks, will have the greatest impact in improving the network variables. Critical blocks in the study area are shown in the following figure.

Blocks with critical power losses, and strong dependence between power losses and LV feeder length

Blocks with critical power losses, and strong dependence between power losses and trans installed cap

Blocks with critical power losses, and strong dependence between power losses and connected load

Blocks with critical voltage drop, and strong dependence between voltage drop and LV feeder length

Fig 7: Critical areas identification using Clustering analysis of different parameters

Prioritization of corrective actions:

As the results show, the Variables in different blocks are in the critical range, because of different reason; therefore, the appropriate action to modify the network parameters for remove a variable from the critical range cannot be the same in all blocks; since number of critical variables and parameters, varies in different blocks, according to table 2 corrective action on the blocks with the most number of critical variables and parameters are priorotized.

Table2: Non-deterministic Expressions describing the variable and parameters in blocks with critical power loss

block number	clustering result				planning priorities
	voltage drop	cennected load	Trans installed cap	LV feeder length	
22429	very high	high	high	very high	2
22436	high	moderate	low	very high	3
22460	very high	very high	very high	high	1
22455	very high	very high	very high	high	1
22409	high	very high	very high	high	2

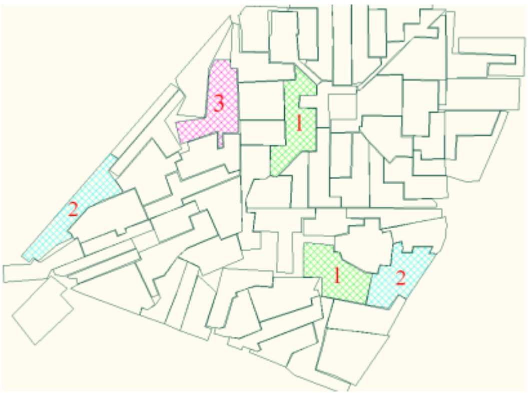


Fig 8: Priority blocks, to modify the parameters specified in Table1

CONCLUSIONS:

An investment optimization is one of the problems in distribution network planning. According to the results of this paper, After determining the critical blocks and identify the most effective parameter in blocks for remove variable from the critical range, modifying these parameters will be investment priorities. In this study area, the first priority to reduce losses and voltage drops is transferring load from 22460 and 22455 blocks to adjacent blocks and increasing transformer installed capacity in these blocks.

REFERENCES

[1] G.Cartina, G.Grigoras, M.Istrate, 2011, "Clustering Techniques for Energy Losses Evaluation in Distribution Networks ", *International journal of mathematical models and methods in applied* science, June 28th - July 2nd, Vol.5, 2011, 133-141

[2] E.C.Bobric, G. Cartina, G. Grigoras, 2009, "Fuzzy Techniques used for Energy Losses Determination in Medium and Low Voltage Networks", *Journal of Electronics and Electrical Engineering*, vol. 90, 2009, 95-98

[3] I. Yatskiv,L. Gusarova, 2005, "The Methods of Cluster Analysis Results Validation", *Proceedings of International Conference RelStat'04*, 2005, 75-80

[4] E.BOBRIC, G.CARTINA 2009, "Clustering Techniques in Load Profile Analysis for Distribution Stations ", *Advances in Electrical and Computer Engineering.* vol. 9, 63-66.

[5] G.Gan, C.Ma, J.Wu 2007, *Data Clustering Theory, Algorithms, and Applications*, 1nd ed. ASA & SIAM, Virginia.

[6] A.D. Gordon, 1999, *Classification*, 2nd ed. Chapman & Hall, NewYork.

[7] M. Momeni, 2011, *Data Clustering*, 1nd ed. Tehran.

[8] N. Hussein, 2003, *A Fast Greedy k-means Algorithm, Master's thesis, Computer Science, University of Amsterdam, Amsterdam, Netherlands.*