

Micro-Spatial Electricity Load Forecasting Using Clustering Technique

1st Christine Widyastuti

Electrical Engineering Departement
Institut Teknologi-PLN
Jakarta, Indonesia
christine.widyastuti@itpln.ac.id

2nd Adri Senen

Electrical Engineering Departement
Institut Teknologi-PLN
Jakarta, Indonesia
adrisenen@itpln.ac.id

3rd Oktaria Handayani

Electrical Engineering Departement
Institut Teknologi-PLN
Jakarta, Indonesia
oktaria@itpln.ac.id

Abstract—Low growth of electricity load forecast eliminates cost opportunity of electricity sale due to unserviceable load demands. Meanwhile, if it is exorbitant, it will cause over-investment and incriminate investment cost. Existing method of sector load is simplified and easy to implement. However, the accuracy tends to bias over one area of which data is limited and dynamic service area. Besides, the results of its forecast is macro-based, which means it is unable to show load centers in micro grids and failed to locate the distribution station. Therefore, we need micro-spatial load forecasting. By using micro-spatial load forecast, the extrapolated areas are grouped into grids. Clustering analysis is used for grouping the grids. It generates similarity matrix of similar data group. Clustering involves factors causing load growth at each grid; geography, demography, socio- economic, and electricity load per sector. Results of every cluster consist of different regional characteristics, which later the load growth is projected as to obtain more accurate forecast.

Index Terms—Forecast, Micro-spatial, Grid, Cluster.

I. INTRODUCTION

ELECTRICITY power supply chain - generator, transmission and distribution started from electricity power supply panning. Electric utility forecast is one pivotal step in planning of electricity development. Forecasting is employed through anticipation model to fulfill electricity needs of electricity supply at certain period of time in an observed area [1]. Load demand depends on different exogenous factors like temperature, humidity, wind speed, seasonal patterns related to human activities and cyclic information [15].

In general, forecasting method is macro based model. As consequence, it is unable to show load centers in micro grids and failed to locate the distribution station. Thus, macro model for forecasting cannot be applied in creating master plan of electricity distribution. For this reason, micro-spatial load forecasting needs to be implemented in micro grids.

Micro-spatial load forecasting methods fall into two categories; trending and land use simulation. Multivariate simulation analysis is defined as a technique to simulate interactions of factors causing loads growth using historical data and condition of existing area by setting the area into certain grids.[9]. This technique has no issue in the data availability at the onset. However, data availability becomes a problem in developing country and is worsen by the fast demands of electricity supply for improving financial level of the area.

Clustering is the process of grouping a set of physical or abstract objects into classes of similar objects. Clustering analysis is a multivariate technique which aims to categorize objects based on their characteristics and solves the issue on large calculation process.

II. MICRO-SPATIAL LOAD FORECAST

A. Time Series Method

Time series can provide further information on how subject change. How to classify accurately has become a hot research point since it is an important element in many computer vision and pattern recognition applications [11]. Micro-spatial load forecast using trending is exploration of historical data and past data growth to predict future load growth.

Baxter (2001) defined time series as a set of observations at sequence of time. Time series which is indexed as Z_t conceptually is interpreted as random variable. Sequence Z_1, Z_2, \dots or \dots, Z_0, Z_1, \dots is generated from stochastic process defined on probability mechanism, which is formulated in the following function:

$$Z_t = F_t + e_t \quad (1)$$

Here, we define Z_t as observation of time series, F_t as a set component of trend, cyclic, seasonality and statistics, and e_t as error.

B. Gompertz Method

Mathematical model of load forecasting using Gompertz method takes growth feature of electricity demands in certain area, where the area in smaller space tends to form unexpected ascending line and reach saturation. Willis called this a progression following curve 'S', in which asymmetric to its points of inflection. Mathematical growth curve according to (Draper and Smith, 1998) is denoted in the following equation:

$$\omega = \alpha \exp\{-\beta e^{-kt}\} \quad (2)$$

with:

α : Asymptotic at value t approaching infinity

β : Integration constant

k : Growth rate

ω : Body weight at time

t : age (days).

C. Multivariate Simulation Method

The diverse output of each independent predictor provides an opportunity to enhance the overall forecast by exploring all viable solutions. Multiple predictors organized in a systematic way called ensemble network can provide better prediction results as compared to a single predictor [6].

Spatial load prediction using multivariable analysis is a method of projecting load type and density of area development based on changes of existing and future land use.

III. CLUSTERING FOR MICRO-SPATIAL LOAD FORECASTING

A. Clustering

Clustering aims to classify grids into relatively homogeneous group of similar characteristics. Hence, load forecast is implemented in a grid model of its cluster. Cluster similarity in clustering process is measured by distance between object vectors (Euclidean distance). The smaller the Euclidean distance between objects, the greater their similarity.[6] Euclidean distance matrix is an $N \times N$ matrix representing the space between N objects.

$$D = \{d_{ij}\}; i, j = 1, 2, 3, 4, \dots, N \quad (3)$$

The commonly used method of distance revision is Euclidean distance representing distance between objects.

$$= \sqrt{\sum_{k=1}^n (v_{ik} - v_{jk})^2} \quad (4)$$

With d_{ij} : Euclidean distance

v_{ik}, v_{jk} : Score grid -i and -j on variable-k

The lowest space euclidean show that grid relation is closely. Repeat step a and b counted $(N - 1)$ numbers, until all of object stay in one group. Write every results of clustering C_j . Grouping result and grouping advantage (C_j) are displayed in dendrogram. Dendrogram can determine total of cluster and its members. Clustering process attempts to minimize objective function, set in clustering process. The goal is generally to minimize within cluster variance and maximize between clusters variance. In other words, grouping data of similar characteristics into one cluster and grouping data of dissimilar characteristics into other group

B. Principal Component Analysis (PCA)

PCA is used to preserve important variables of each cluster obtained in clustering process. PCA depends on the type of original dataset. If original variables have the same units, the principle component is derived from covariance matrix. Whereas when original variables have different units of measurement, the principle component is derived from correlation matrix R . Thus, this needs transformation of original data into standard form:

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (5)$$

With \bar{x}_j : variable means -j

s_j : variable standard deviation -j

Total of principle components are based on the proportion of cumulative variance of 75% or more than total variance.

C. Factor Analysis

Factor analysis used to describe the structure of observed variables by generating factors in lower number than the observed variables. It consists of two analyses; component analysis and common factor analysis. The score of correlation coefficient between original variables and principle component variables is written as a function as follows:

$$r_{ij} = \alpha_{ij} \sqrt{\lambda_j} \quad (6)$$

Where r_{ij} : Correlation coefficient among variables

λ_j : Total variance explained

D. Mathematical Model

The model is formulated in multiple regressions based on the following mathematical model [2]

$$Y = b_1 + b_2 X_2 + b_3 X_3 + \dots + b_k X_k + e \quad (7)$$

Simplified into matrix:

$$Y = Xb + e \quad (8)$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_k \end{bmatrix} \quad e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_n \end{bmatrix}$$

Y : load density matrix $n \times 1$

X : variables matrix $n \times k$

B : regression coefficient matrix $k \times 1$

e : error matrix $k \times 1$

To obtain the values of b , the number of squared deviation is minimized:

$$\sum e_i^2 = e'e = (Y - Xb)'(Y - Xb) \quad (9)$$

E. Tren Variables

To obtain load density growth at each year based on previous load density, we need to calculate trend of respective variable (except land use variables). The most appropriate trend variable to select is based on the lowest mean absolute percentage error (MAPE).[10]

Trend of land use change is taken from spatial planning and regional planning of involved area. With environmental the data of neighborhood and hamlet were range between the next 5-10 years, we refer to trend of regional development in past years as to determine land use change.

F. Peak Load Forecast

After trend of each variable is obtained (except for land use using RT/RW and historical data), variables trend growth are used for forecasting the load density following past load density model.

Results of load density forecast at each year are obtained from this cluster. Further, the results are used to calculate load density of respective sector of similar cluster. With the density of each sector, we can estimate the capacity of power by sector and district. This is obtained by multiplying load density per sector with the length of sector of its district in the cluster. Meanwhile, changes of length of sector at each year is adjusted to spatial regional planning. Further process is to calculate total power of district by summing total energy per sector (housing, business, industry and social), which is formulated in the following equation:

$$P_{Totalofneighborhood}(t) = C_f(P_R(t) + P_B(t) + P_I(t) + P_S(t)) \quad (10)$$

where C_f : Coincident factor

G. Flowchart of microspatial load forecast

Here is flowchart of all methods, displayed as follows:

IV. RESULT AND DISCUSSION

A. Data

Clustering process involves a wide range of electricity variables and non-electricity variables. Total of variables are 12, and total of districts are 114. However, here we only provide 10 districts as seen in the table below:

B. Building Cluster

Area taken as the object of research is area network including some parts of Tangerang, Bogor and West Jakarta. There are 114 districts taken as grids for clustering. According to data mining, two methods of clustering are hierarchical clustering and non-hierarchical clustering [6]. Hierarchical clustering is algorithm that group two or more similar objects into cluster. The process is repeatedly executed until clusters merged together producing tree-like diagram, which shows the hierarchical relationship between objects. Dendrogram is usually the output to describe the hierarchical process. [8].

Clustering is employed by grouping objects (districts) into clusters, in which every cluster consists of district with relatively homogenous characteristics. Objects grouping are implemented by clustering technique.

Agglomerative clustering starts from N cluster to singleton cluster, in which N is the total of data, while divisive clustering starts from singleton cluster into N cluster. The process can be seen in the figure below:

From results of clustering, the total of 114 districts is grouped into 5 clusters, as seen below:

C. Cluster Characteristics

When clustering is applied, every cluster has relatively homogenous characteristics for each district, as depicted below:

Hierarchical clustering is able to find the similar data into one hierarchy and dissimilar data into another group. To explain the distinctions of characteristics' illustration of respective cluster at each dimension, we use discriminant analysis. Procedure started after cluster is determined.

D. Calculation Of Load Growth Of Each District

Based on the results, trend of each variable (except for land use using RT/RW) is used to forecast load density. Further process is to calculate total power of districts by summing total energy per sector (housing, business, industry and social).

Results of energy growth are elaborated from load per sector, which demands in industrial sector is higher, followed by business sector of 562.2 MVA and housing sector. Meanwhile, social sector has the lowest energy demands. We can see that average percentage of industrial growth is more dynamic compared to other sector loads. System growth increases with linear raise with the average of 6.39 %. Average growth in other sectors; housing 7.05 %, industry 6.63 %, business 6.0 %, and social 6.93 %. On the final note, it shows that the average growth of each sector is almost similar, that is around 6 %.

V. CONCLUSION

Micro-spatial load forecasting is able to identify and classify area into smaller group of homogenous characteristics from each district by clustering technique. Using clustering, 114 districts are reduced into 5 clusters. Results of forecasting is an elaboration of macro-spatial load forecasting which provides solution such as information of load size, time and location with higher accuracy level. For this reason, this method is appropriate for basic planning of electricity network distribution development.

The accuracy level of this method will be better if the extrapolated area is smaller and the data validity provided is accurate. As the basis of load variables are load density, this method can accommodate immediate load changes based on allocation of spatial based on sector service.

Due to its easy programming system, this technique is feasible to be perfected by enhancing its ability in pattern recognition through fuzzy application or other intelligent systems.

REFERENCE

- [1] Mujiati Dwi Kartikasari and Arif Rohmad Prayogio(2018) "Demand Forecasting of Electricity in Indonesia with Limited Historical", J. Phys.: Conf Data.
- [2] X. Sun, Z. Ouyang and D. Yue (2017) "Short-Term Load Forecasting Based on Multivariate Linear Regression," IEEE Conference on Energy Internet and Energy System Integration (EI2), Beijing.
- [3] JP Carvallo, PH Larsen, AH Sanstad, CA Goldman (2017), "Load Forecasting in Electric Utility Integrated Resource Planning," *osti.gov* 6.

- [4] Laurinec, Peter Lucka, Maria.(2017) “New clustering-based forecasting method for disaggregated end-consumer electricity load using smart grid data”. 210-215. 10.1109/INFORMATICS.8327248.
- [5] Krzysztof Gajowniczek, Tomasz Zabkowski (2018) “Simulation Study on Clustering Approaches For Short-Term Electricity Forecasting”, Complexity, Complex Optimization and Simulation in Power System.
- [6] Raza, Muhammad Qamar Nadarajah, Mithulananthan Li, Jiaming Lee, Kwang, ”Multivariate Ensemble Forecast Framework for Demand Prediction of Anomalous Days”. IEEE Transactions on Sustainable Energy, p. 6, 2018
- [7] Jianwei Mi, Libin Fan, Xuechao Duan , and Yuanying Qiu (2018) “Short-Term Power Load Forecasting Method Based on”, Hindawi, Mathematical Problems in Engineering. Improved Exponential Smoothing Grey Model2017.
- [8] Shahzadeh, Abbas Khosravi, Abbas Nahavandi, Saeid (2015) “Improving load forecast accuracy by clustering consumers using smart meter data”. 1-7. 10.1109/IJCNN.7280393.
- [9] A Senen (2013), “Study of Microspatial Electricity Load Forecast Based on Land Use Simulation” Journal of Electrical Media.]
- [10] Adri Senen, Titi Ratnasari (20116) “Forecasting Study of Short-Term Average Load Using the Autoregressive Integrated Moving Average (ARIMA) Method ”, SUTET Scientific Journal, Vol. 7, No.2.
- [3]
- [3]
- [3]
- [3]