Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Simple Estimation of Operating Lifetime for Electrical Distribution Components Utilizing Their Inspection and Maintenance Records

Hirotaka Takano*. Kan Nakae*. Tatsuya Tokunaga. Toshihiro Hayashi****

*Department of Electrical, Electronic and Computer Engineering, Gifu University,
Gifu, 501-1193 JAPAN (e-mail: takano@gifu-u.ac.jp)
**Kansai Transmission and Distribution, Inc., Osaka, 530-8270 JAPAN

Abstract: Distribution system operators (DSOs) inspect electrical components in distribution systems and decide necessary actions for the inspection results to sustain the reliability of power supply. These results have been accumulated in databases as the inspection and the maintenance records. The authors propose to utilize the inspection and the maintenance records in estimation of operating lifetime of newly constructed electrical components in the distribution systems. A decision tree learning analyzes the relationship between attributes of the target components and the operating lifetime obtained by the inspection and the maintenance records, and as a result, constructs a tree-like classification or regression model. The resulting model can estimate the operating lifetime of newly constructed components using the initial information (the set of the attributes) only. Usefulness of the estimation method is verified on actual records stored in a Japanese electric power company. In addition, factors that have strong influence on the operating lifetime are specified by the feature of decision tree.

Keywords: distribution equipment, lifetime estimation, decision trees, inspection record, maintenance record, big data

1. INTRODUCTION

Electrical power distribution systems consisting of distribution feeders, electric poles, pole transformers and switchgears have an extremely important role on delivering electricity from suppliers to consumers. If a trouble occurs in the distribution systems, the resulting outage, as is well known, brings significant impacts on activities of our society. With a view to sustaining the stable power supply, distribution system operators (DSOs) have been making inspection and maintenance for electrical components sequentially (Department of Defense. 2017). Through the inspection and the maintenance, conditions of the electrical components (sets of the checklist and the scores in them) and necessary actions for the inspection results, e.g. follow-up observation, repair or replacement, are accumulated in the DSOs.

These records have been growing rapidly associating with the progress of automation and digitalization. Since gathered information often includes useful rules, knowledge and judgement criteria (Segaran, T. and Hammeribacher, J. 2009), we can expect to improve the reliability, the quality and the efficiency in operations and planning of the distribution systems by its appropriate applications (Investigating R & D Committee on Energy-Related Big Data and their Applications. 2019; Takano, H. et al. 2019). However, applications of the records have been normally limited to utilization in the decision-making whether the electrical components require to be taken measures or not for keeping their functions (Khuntia, R.S. et al. 2016). This is because the total numbers of the distribution system components are indeed large (Short, T.A. 2004; Takano, H. et al. 2016), and therefore, it is difficult to analyze and utilize these records effectively relying on knowledge and experience of the DSOs only. That is, there is still plenty room for discussion on how to utilize the massive records, and what kinds of techniques are suitable for analyzing and utilizing them (Barriquello, C.H. 2017; Shimasaki, H. et al. 2019; Shiomi, R. et al. 2019).

The authors propose an estimation method for operating lifetime of newly constructed electrical components in the distribution systems as an application of the inspection and the maintenance records. Now, the authors define the operating lifetime as the period from the operation starting date to the date when the operators required "replacement", that is available period of the target component. There are three steps in the proposed estimation method. First, a decision tree learning is applied to analyze the relationship between attributes of the target components, e.g. type, size and location, and their operating lifetime. The attributes are included in the inspection record, and the operating lifetime can be calculated by referring both records. Next, a tree-like estimation model is constructed as a result of the analysis. In this step, the constructed tree model can visually provide factors having influences on the operating lifetime, and this is the strongest reason why the authors employed the decision tree learning. Finally, the constructed model estimates the operating lifetime of newly constructed electrical component using its initial information (set of attributes) as the input. Since the initial information is easily available for the DSOs, this method can be applied in the distribution system planning stage.

Usefulness of the aforementioned proposal is verified through numerical simulations and discussions on their results. In the numerical simulations, the inspection and the maintenance records actually stored in a Japanese electric power company are used. Factors (sets of the attributes and their condition) that have strong influence on the operating lifetime are also specified as a result of the proposed method.

2. ESTIMATION CONDITION

In the distribution system planning stage, the DSOs make several planning decisions with the limited information. This is the reason why the authors focused on attributes of the electrical components in the inspection record. The attributes are roughly classified into the specification and the locational condition. For example, in the concrete electric poles, type and length are available for the specification attributes, and soil quality and salt damage level are in the locational attributes. This section introduces an overview of the attributes and sets the average operating lifetime in the target datasets.

Although the authors applied the proposed method to three components in the distribution systems (concrete electric poles, pole transformers and switchgears), the concrete pole, which has the largest share in available datasets, is emphasized as a typical distribution component in this paper. Table 1 shows comparison of total numbers of datasets of each component, and Table 2 summarizes the breakdown of the concrete poles.

Table 1. Comparison of total numbers of components

Concrete pole	Pole transformer	Switchgear
1,382,067	1,056,191	95,776

Table 2. Breakdown of target data (concrete pole)

Total number of all concrete poles	1,382,067
Total number of concrete poles including missing data (removed from discussions)	28,143
Total number of concrete poles finishing their lifetime (available for model construction)	4,951
Total number of concrete poles having lifetime under 20 years (removed from discussions)	266

2.1 Attributes of Target Component

In the analysis step, the relationship between the attributes of the target components and the actions for the inspection results (follow-up observation, repair and replacement) are analyzed. Moreover, the attributes are utilized as the input of the estimation step. Table 3 shows the attributes of the concrete pole. Ellipsis is for simplifying the estimation model.

As shown in Table 3, ten attributes are selected as the initial information for the concrete poles. The specification attributes consist of type (TYP), length (LEN), facility type (FAC), and with or without pole transformer (TRA) or switchgear (SWI). In contrast, the locational attributes include salt damage level (SAL), surrounding condition (SUR), geological condition (GEO), soil quality (SOI) and total number of concrete poles existing in same area (NUM). The inspection record also has scores or actual values corresponding with each attribute. Since the proposed method is aiming to apply in the planning stage, the inspection results (condition of the target components) are not utilized in this study.

Table 3. Attributes of concrete poles

Input attributes	Ellipsis	Contents
Type	TYP	Symbol A to Z
Length	LEN	8 to 30 (m)
Salt damage level	SAL	Score 1 to 3
Surrounding condition	SUR	Score 1 to 7
Geological condition	GEO	Score 1 to 3
Soil quality	SOI	Score 1 to 5
Facility type	FAC	Score 1 to 4
Total **num**ber of concrete poles existing in the same area	NUM	10,694 to 60,172 (poles)
Transformer (w or w/o)	TRA	1 (w) or 0 (w/o)
Switchgear (w or w/o)	SWI	1 (w) or 0 (w/o)

In the case that we change the estimation target, five specification attributes (TYP, LEN, FAC, TRA and SWI) are replaced with those of the other components.

2.2 Calculation of Average lifetime

With a view to setting the basis of discussion, the average operating lifetime of the concrete poles is calculated. According to the definition in Section 1, the average lifetime can be obtained by associating the operation starting date and the date when the operators described “replacement”. Although 4,951 concrete poles are available for the calculation, the authors removed 266 concrete poles having too short lifetime (under 20 years) as the samples of irregular replacement (e.g. disaster-originated replacement). Figure 1 displays the distribution of operating lifetime of the target concrete poles and their average lifetime.

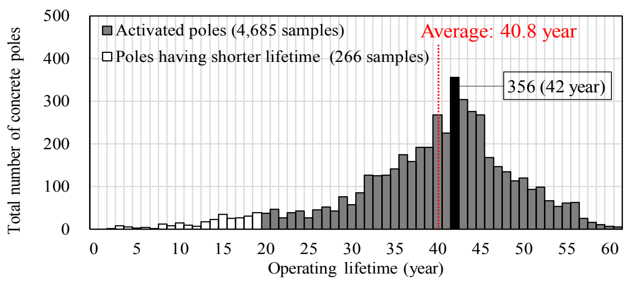


Fig. 1. Distribution of operating lifetime (concrete pole).

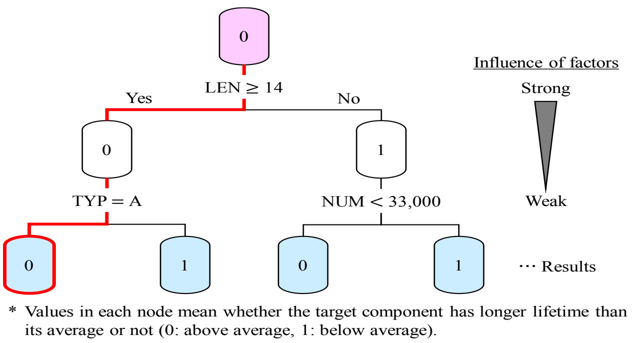
In Fig. 1, the average lifetime of the concrete poles is calculated as 40.8 year. Generally, in Japan, the average lifetime of concrete poles has been estimated in the range of 30 to 40 years by knowledge and experience of the DSOs. Besides, it is well known to change the actual lifetime depending on various factors (Miyagawa, T. 1991; Kakizaki, M. 1991). From Fig. 1, it can be confirmed that the target concrete poles have longer lifetime than their general average.

3. ESTIMATION METHOD

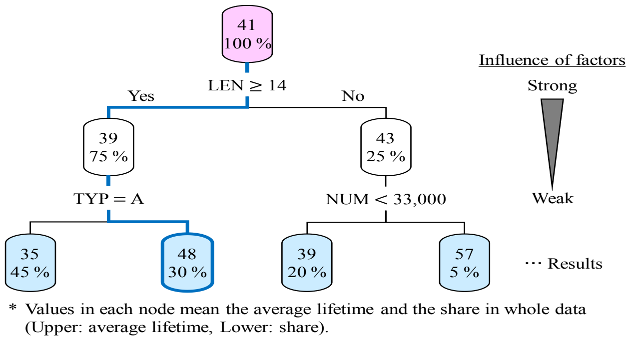
There are various estimation models in machine learning techniques including multiple regression models, artificial neural networks and decision tree models (Rawlings, J. et al. 2001; Menzies, T. and Hu, Y. 2003; Segaran, T. 2008; Haykin, S. 2008; Rokach, L. and Maimon, O. 2014). In this paper, a decision tree learning is selected as the basis of the operating lifetime estimation. The decision tree learning constructs a tree-like model representing its decisions and decision-making process visually and explicitly, and therefore, we can easily understand the judgement criteria without any special knowledge and experience for its algorithm. The authors emphasized this feature to specify the factors having strong influence on the operating lifetime. This section describes details of the proposed estimation method.

3.1 Outline of Lifetime Estimation

Decision tree learning is a knowledge representation that ultimately makes some kind of decision by accumulating questions about the attributes of objects. The decision tree is classified into the classification and the regression models. In this paper, both tree models are constructed. The former estimates whether the operating lifetime of the target exceeds the average or not (tendency in the operating lifetime). Meanwhile, the latter is used for estimating value of the operating lifetime (available period). Figure 2 illustrates samples of the classification and the regression tree models.



a. Classification tree.



b. Regression tree.

Fig. 2. Sample of constructed decision tree models.

As illustrated in Fig. 2, the decision tree model normally consists of nodes, branches and leaves. The node represents a question on an attribute, and each leaf node expresses a class. The paths from root to leaf represent classification rules. In Fig. 2a, the classification tree judges that the target component has shorter lifetime (class in the classification tree) through the answers for questions, “Length is longer than or equal to 14 m (LEN ≥ 14)” and “Type is A (TYP = A)”. On the other hand, in Fig. 2b, the regression tree estimates that the operating lifetime of the target component is 48 year (class in the regression tree) through the answers for questions, “Type is not A” and “Length is shorter than 14 m”. It also means that 30 % of all datasets included in the model construction step had the same characteristics.

3.2 Outline of Algorithm

Typical approaches for creating the decision tree include Classification and Regression Tree (CART), Iterative Dischotomiser 3 (ID3) and C 4.5 (Quinlan, J.R. 1993; Singh, S. and Giri, M. 2014). The authors select a CART algorithm, which is one of the most popular approaches for non-parametric decision tree learning. The CART model involves selecting input variables and split points on those variables until satisfying the convergence criterion. This process is called “growth”.

In this study, Gini impurity is used as the splitting criterion in the classification tree. When the Gini impurity is taken as the branch criterion, the attribute, which reduces most the Gini impurity after the branching operations, is selected as the question in the node (Kaneda, S. 1991; Loh, W. 2011). The Gini impurity is defined as

$$GINI_n = 1 - \sum_{i=1}^{CL} p(i|n)^2 \ (= 1 - \sum_{i=1}^{CL} \left(\frac{S_i}{S_n}\right)^2), \qquad (1)$$

where n is the node number ($n = 1, 2 \cdots, ND$); i is the class number (need replacement or not) ($i = 1, 2 \cdots, CL$); $p(i\,|\,n)$ is the probability density function; S_n is the total number of sample data in Node t; S_i is the total number of sample data included in Class i.

Meanwhile, in the regression tree, the residual sum of squares (RSS) is employed. The RSS is used to measure the amount of variance in a dataset and expressed with

$$RSS_n = \sum_{j=1}^{SD} \left(y_j - \mu[n]\right)^2, \qquad (2)$$

where j is the number of sample data ($j = 1, 2 \cdots, SD$); y_j is the actual operating lifetime of Sample j; $\mu[n]$ is the average lifetime in Node n.

Since the resulting maximum tree models have the possibility of overfitting exists, verbose paths are integrated until the following complexity cost becomes sufficiently small. This process is called “pruning”.

$$E = \sum_{n \in LN} \left(\frac{S_n}{S} \cdot F_n\right) + \alpha \cdot ND, \qquad (3)$$

where F_n is the Gini impurity or the RSS in Node n; LN is the set of leaf nodes; S is the total number of all sample data; α is the control parameter.

4. NUMERICAL SIMULATION

In order to verify the usefulness of the authors’ proposal, numerical simulations were carried out using the records of the inspection and the maintenance described in Section 2. As shown in Table 2, there are 4,685 available electric poles, of which 4,451 (95 %) datasets were used in steps of the data analysis and the model construction (training process). The remaining 234 (5 %) datasets were used in the estimation step (verification process) as the datasets of newly constructed concrete poles. While interchanging the verification datasets with the learning datasets, the estimation models were repeatedly constructed 20 times. Table 4 shows the dataset used in the training process, and Table 5 summarizes the input datasets in the verification process.

Table 4. Data for training process

	Below average	Above average	Total
Training dataset	1,969	2,482	**4,451**

Table 5. Input data for verification process

	Below average	Above average	Total
Training dataset	1,969	2,482	**4,451**
Verification dataset	106	128	**234**
Good samples	–	286,137	**286,137**

In the application of classification tree (tendency estimation), the operating concrete poles already exceeding the average lifetime were added to the estimation targets. These concrete poles were described "good samples" in Table 5 (286,137 poles). In contrast, in the regression tree (duration estimation), only the aforementioned 4,685 concrete poles were used for the verification process. Now, parameters in the estimation were set by trial and error, and discussions on their appropriateness remain as an important issue of this study.

With the aim of specifying the factors (sets of the attributes and their condition) having the strong influence on the operating lifetime, the following index is defined (Sugimoto, T. et al. 2007) and applied to the constructed tree models.

$$V = \sum_{n \in BR} \Delta F_n(x_k), \tag{4}$$

where $\Delta F_n(x_k)$ is the mean decrease in the Gini impurity or the RSS for the attribute x_k; BR is the set of nodes excluding leaf nodes.

4.1 Results for Classification Tree

Table 6 summarizes results of the tendency estimation for each input dataset. Figure 3 illustrates the best classification tree in 20 cases, and Fig. 4 shows analysis results of the judgement criterion in Fig. 3.

Table 6. Estimation accuracy in classification tree

Estimation target		Below average	Above average	Overall average
Training datasets	Best	81.3 %	72.4 %	75.0 %
	Average	79.4 %	68.3 %	73.2 %
	Worst	77.5 %	64.7 %	71.2 %
Verification datasets	Best	82.1 %	73.7 %	74.9 %
	Average	76.0 %	65.4 %	70.1 %
	Worst	71.2 %	58.2 %	66.7 %
Good samples	Best	–	73.1 %	73.1 %
	Average	–	70.5 %	70.5 %
	Worst	–	68.7 %	68.7 %

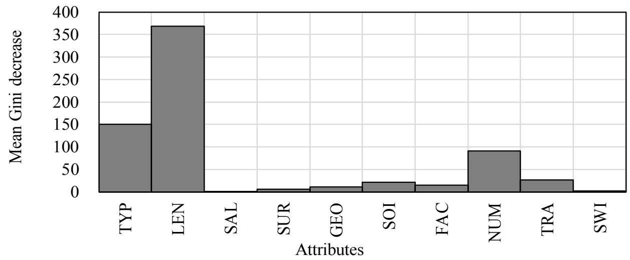


Fig. 4. Analysis result in classification trees (average).

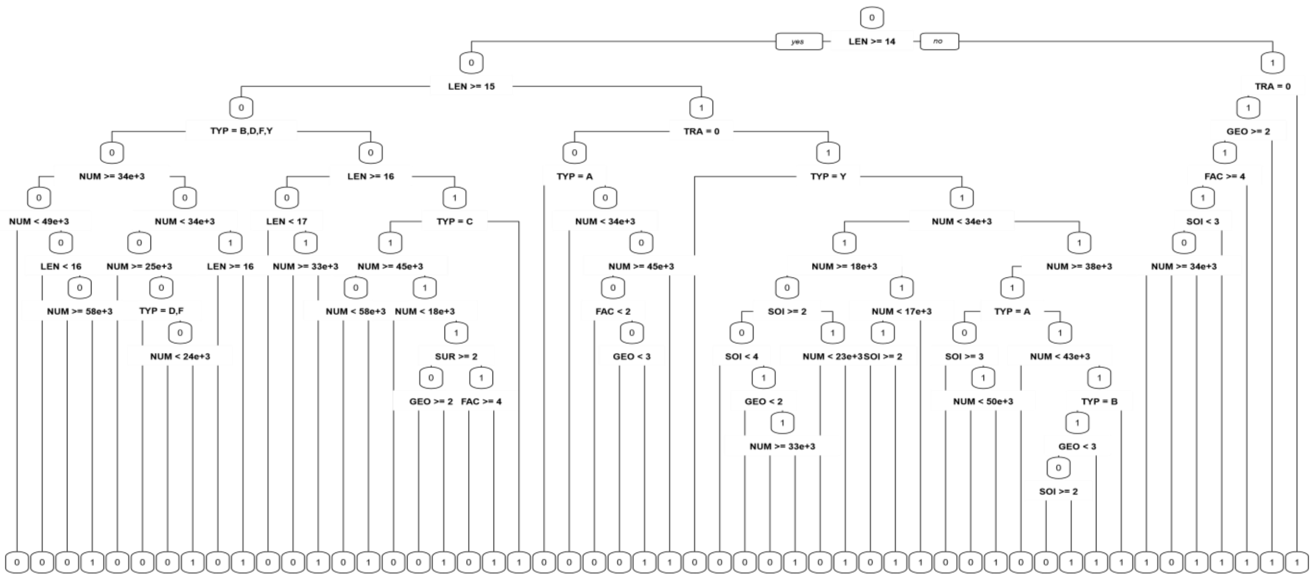


Fig. 3. Constructed classification tree (best result).

As shown in Table 6, the average values in the estimation accuracy exceeded 70 % in each dataset even though the proposed estimation method used the initial information only. In addition, there was no significant difference between the average and the best or the worst cases in both the training and the verification datasets. From these results, the authors concluded that the decision tree learning functioned well. Meanwhile, in Fig. 3, length (LEN), type (TYP) and total number of concrete poles existing same area (NUM) were frequently selected in the constructed classification tree. By Fig. 4, it can also be confirmed that these attributes were importance in the classification tree.

4.2 Results for Regression Tree

Results of the duration estimation is summarized in Table 7. Figure 5 illustrates the best regression tree in 20 cases, and Fig. 6 displays analysis results of the judgement criterion. In Table 7, the root-mean-square-error (RMSE) was calculated as

$$RMSE = \sqrt{\frac{1}{L}\sum_{l=1}^{L}(z_l^* - z_l)^2}, \qquad (5)$$

where l is the number of input concrete pole ($l = 1, 2, \cdots, L$); z_l^* is the actual operating lifetime of the input data l; z_l is the estimated operating lifetime of the input data l.

As shown in Table 7, values of the estimation error were approximately 15 %, and most of the RMSE values were lower than 7 years. Moreover, there was no significant difference between the average and the best or the worst cases in both the training and the verification datasets. Since the average lifetime in the datasets was 40.8 year as described in Section 2, we can expect that the proposed method contributes to improve the estimation accuracy in the distribution system planning. In Fig. 5, length (LEN), type (TYP) and total number of concrete poles existing same area (NUM) were selected frequently by the constructed regression tree similar to the classification results. With reference to Fig. 6, we can understand that these attributes had very strong influence on the operating lifetime as compared to the others.

Table 7. Estimation accuracy in regression tree

Estimation target		Estimation error	RMSE
Training datasets	Best	13.5 %	6.5 year
	Average	13.7 %	6.6 year
	Worst	14.0 %	6.7 year
Verification datasets	Best	12.4 %	6.0 year
	Average	14.3 %	6.8 year
	Worst	15.7 %	7.3 year

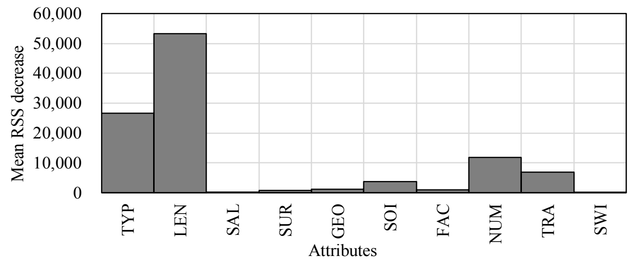


Fig. 6. Analysis result in regression trees (average).

4.3 Discussions

In the estimation results, the operating lifetime of concrete poles strongly depended on their length in both the classification and the regression trees. These results were verified with reference to the information described in the inspection and the maintenance records. Figure 7 displays the relationship between the operating lifetime and the length of concrete poles, and Fig. 8 shows the distribution of pole length.

In Fig. 7, the operating lifetime decreased gradually in contrast to the length of concrete pole until the length reaches 17 m. Besides, as shown in Fig. 8, most of the concrete poles have length from 12 m to 16 m. These results show the possibility that the length of concrete poles had actually influence on their operating lifetime. Furthermore, the authors examined for the type and the total number and confirmed the similar tendency in them.

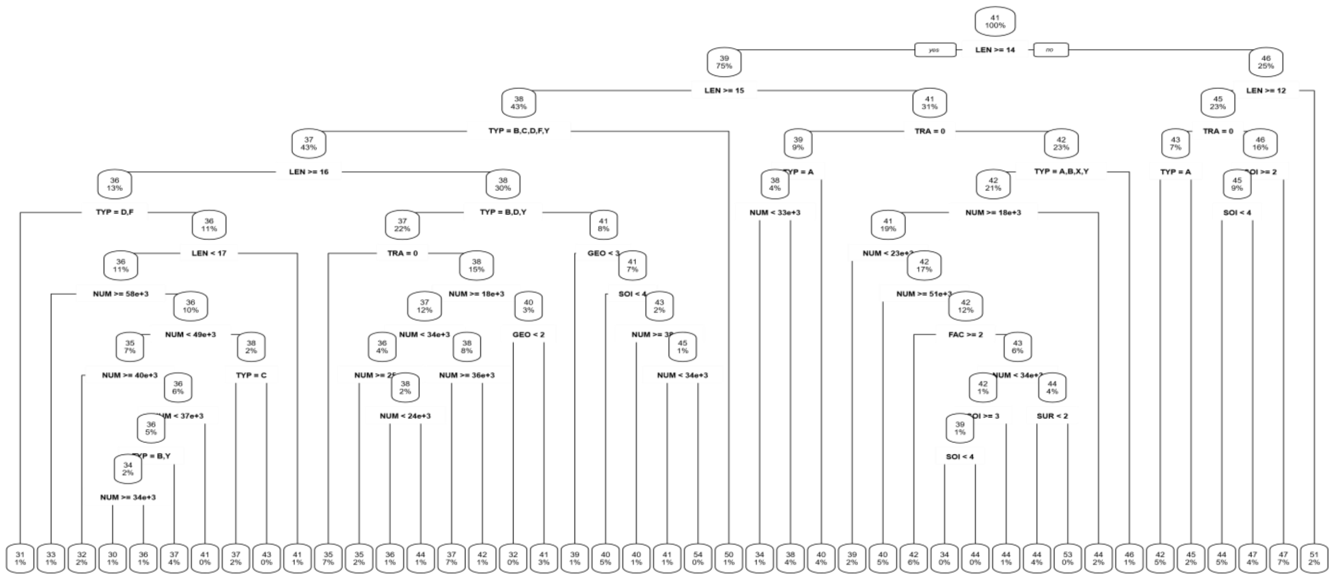


Fig. 5. Constructed regression tree (best result).

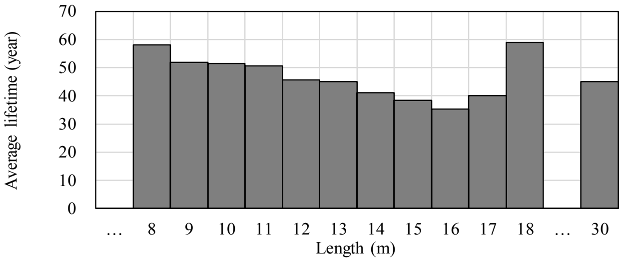


Fig. 7. Average lifetime of concrete poles on each length.

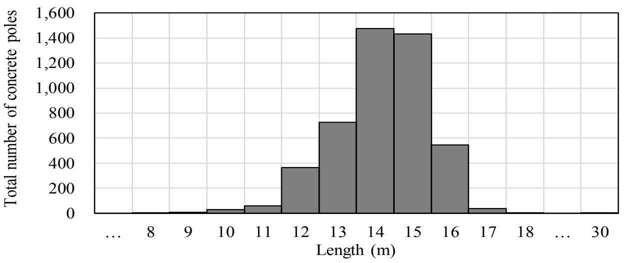


Fig. 8. Length distribution of concrete poles.

5. CONCLUSIONS

This paper presented an estimation method for operating lifetime of newly constructed electrical distribution. Through the numerical simulations and the discussions on their results, the authors concluded that the proposed method can support the decision-making in the distribution system planning. In future works, the authors will improve the estimation accuracy with respect to optimization in the parameter tuning.

REFERENCES

Barriquello, C.H., Garcia, V.J., Schmitz, M., Bernardon, D.P. and Fonini, J.S. (2017. A Decision Support System for Planning and Operation of Customer Services in Electric Power Systems. In Volosencu, C. *System Reliability*, pp. 355-370. In Tech, Croatia.

Department of Defense. (2017). Operation and Maintenance (O&M) Exterior Power Distribution Systems: Unified Facility Criteria UFC 3-550-07. https://www.wbdg.org/FFC/DOD/UFC/ufc_3_550_07_2017.pdf [Access 31 Oct. 2019]

Haykin, S. (2008). Neural Networks and Learning Machines Third Edition. *Prentice Hall*, USA.

Investigating R & D Committee on Energy-Related Big Data and their Applications. (2019). Energy-related big data and their applications. *IEEJ Technical Report*, 1441. (in Japanese)

Kakizaki, M. (1991). TEKKIN KONKURI-TOZOU TATEMONO NO JUMYOU YOSOKU (Lifetime prediction of reinforced concrete building). *Journal of Japan Society for Safety Engineering*, 30(6), pp. 421-431. (in Japanese)

Kaneda, S. (1999). KETTEIGI GAKUSHU: JITSUYOUKI WO MUKAETA NONPARAMETORIC TOUKEI SHUHOU (Decision tree learning: Nonparametric statistical method that has reached the practical stage)," *CICSJ Bulletin*, 17(4), pp. 9-14. (in Japanese)

Khuntia, R.S., Rueda, J.L., Bouwman, S. and Meijden, A.M. (2016). A literature survey on asset management in electrical power [transmission and distribution] system. *Int. Trans. on Electr. Energ. Syst.*, 26, pp. 2123-2133.

Loh, W. (2011). Classification and regression trees. *WIREs Data Mining and Knowledge Discovery*, 1, pp. 14-23.

Menzies, T. and Hu, Y. (2003). Computing Practices – Data Mining for Very Busy People. *Computer*, 36(11), pp. 22-29.

Miyagawa, T. (1991). DOBOKU KONKURI-TO KOUZOUBUTSU NO JUMYOU YOSOKU (Lifetime prediction of civil engineering structure). *Journal of Japan Society for Safety Engineering*, 30(6), pp. 415-420. (in Japanese)

Quinlan, J.R. (1993). C4.5: Programs for Machine Learning. *Morgan Kaufmann Publishers*, USA.

Rawlings, J., Pantula, S. and Dickey, D. (2001). Applied Regression Analysis: A Research Tool, Second Edition. Springer, Germany.

Rokach, L. and Maimon, O. (2014). Data Mining with Decision Trees 2nd ed. *World Scientific Publishing Co., Inc.*, Singapore.

Segaran, T. (2008). Programming Collective Intelligence. *O'Reilly Media, Inc.*, USA.

Segaran, T. and Hammeribacher, J. (2009). Beautiful Data: Stories Behind Elegant Data Solutions. *O'Reilly Media, Inc.*, USA.

Shimasaki, H., Shiomi, R., Takano, H. and Taoka, H. (2019). A study on maintenance decision support for power grid components using their inspection and maintenance records. *Journal of International Council on Electrical Engineering*, 9(1), pp. 38-44.

Shiomi, R., Shimasaki, H., Takano, H. and Taoka, H. (2019). A study on operating lifetime estimation for electrical components in power grids on the basis of analysis of maintenance records. *Journal of International Council on Electrical Engineering*, 9(1), pp. 45-52.

Short, T.A. (2004). Electric Power Distribution Handbook. *CRC Press*, USA.

Singh, S. and Giri, M. (2014). Comparative Study Id3, Cart and C4.5 Decision Tree Algorithm: A Survey. *International Journal of Advanced Information Science and Technology*. 3(7), pp. 47-52.

Sugimoto, T., Simokawa, T. and Goto, M. (2007). Tree-Structured Approaches and Recent Advances. *Bulletin of the Computational Statistics of Japan*, 18(2), pp.123-164.

Takano, H. Kawamura, K. and Iizaka, T. (2016). Industry Practice and Operational Experience of Key Distribution Applications of Smart Grid Technologies. In Liu, C.C., McAuthur, S., Lee, S.J. *Smart Grid Handbook, 3 Volume Set*, pp. 967-993. Wiley, USA.

Takano, H., Yano, T., Masahiro, S. and Toyoshima, I. (2019). DENRYOKU BUNYA NO BIG DATA TO KATSUYOU JIREI (Big Data and Its Applications in Electric Power Field). *Smart Grid*, pp. 3-8. TAIGA Publishing Co., Ltd., Japan. (in Japanese)