A STUDY FOR CAUSE ESTIMATION OF FAULTS USING STATISTICAL ANALYSIS

Ryota Yamamoto
TEPCO Holdings, Inc. Japan
yamamoto.ryota@tepco.co.jp

Masato Watanabe
TEPCO Power Grid, Inc. Japan
w.masato@tepco.co.jp

Yoshinori Ogihara
TEPCO Holdings, Inc. Japan
ogihara.yoshinori@tepco.co.jp

ABSTRACT

In distribution power systems, these are many lines and equipment. Therefore, a lot of time and work are required to find the fault points when a power failure is occurred in a distribution line. This paper focuses on a cause estimation method of faults to assist finding a fault point. Specifically, this paper proposes an estimation method based on a statistical analysis to improve estimation accuracy, and confirm the effectiveness of the proposed method by verification using measurement data obtained by switches with sensors.

INTRODUCTION

In general, when a power failure occurs in a distribution line, the location of the fault and the equipment that caused a power failure, that is the point of the fault, needs to be determined for early power recovery.

In a current system to determine the fault range, when the system detects a power failure by a protective relay at a substation, the system stops supplying electricity to the distribution line in which the power failure occurs, and narrows down the range of the power failure to some extent by a timed-sequential transmission method. After that, a fault point is specified by patrols and investigation of leakage current by using superimposed voltage.

The steps up to narrowing down the range of the power failure by the timed-sequential transmission method are done automatically by the system and these steps are completed in a matter of minutes. On the other hand, the patrols and the investigation are done by workers, and require several tens of minutes to several hours. In mountainous regions in particular, distribution lines are long and the range to narrow down by the system tends to be larger, so the time needed to find a fault point also tends to become longer.

In this background, the fault localization methods [1] to specify the point where the fault occurred when a power failure occurs and the methods [2]-[5] to estimate a cause of fault have attracted attention, and various methods have been proposed.

On the other hand, many technical issues still remain in response to a wide variety of faults in distribution lines, such as the fault caused by breakdown of underground cables or burnout of a transformer.

An extraction of feature quantity by a waveform analysis to estimate a cause of fault, in particular, has been systemized, but the extraction largely depends on rules of thumb in the process of estimating causes based on the feature quantity. How the rules of thumb should be theoretically evaluated has been an issue.

However, in recent years, switches with sensors, which have various measurement functions, has been introduced in distribution lines and communication infrastructure such as optical lines has been developed. In addition, circumstances that enables smooth recording and accumulation of voltage and current waveform data and application of statistical analysis have been gradually established.

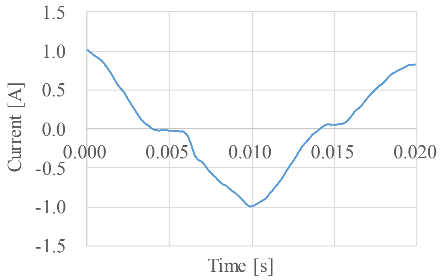
Taking into account such a background, this paper proposes the method to estimate a cause of faults by applying a statistical analysis to a certain number of current waveform data obtained by switches with sensors, without using a collation table of feature quantity and causes of faults, which has been created based on rules of thumb.

EXTRACTION OF FEATURE QUANTITY

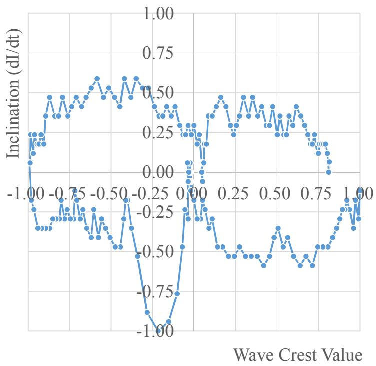
Various methods have been proposed for extracting features of waveforms as values and evaluating them [2]-[5]. Of these methods, this paper introduces the analysis method based on phase-plane trajectory, focusing on evaluating visual shapes of waveforms [2]. This method estimates causes of faults through an extraction of feature quantity from a current waveform (Fig. 1).

First, time derivatives for the current waveform data shown in Fig.1 (a) is obtained at regular intervals. Next, the plot graph called the phase-plane trajectory graph as shown in Fig.1 (b) is made. A horizontal axis shows crest values and a vertical axis shows time derivatives. In order to evaluate the phase-plane trajectory graph numerically, the graph is divided into 64 cells as shown in Fig.1 (c) and the number of points plotted in each cell is counted. Then the number of plots is extracted in a specific cell such as the region A and B in Fig.1 (c) and a scatter diagram is created as shown in Fig.1 (d), which allows you to evaluate the shapes of accident waveforms and estimate the cause.

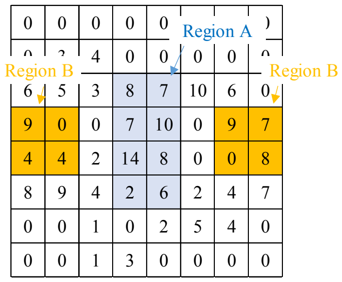
This method has an advantage that visual features of waveform data such as steepness of the waveform gradient, square waves and triangle waves can be numerically expressed easily. On the other hand, in the same way as other techniques, a collation table that links the feature quantity of waveforms and causes of faults needs to be created in advance (for example, when needle-like waves are linked continuously, a cause of a fault is likely to be cable ground fault). For this reason, the improvement of methods to create a collation table has been an important issue, as well as the accumulation of waveform data and the improvement of methods to extract feature quantity in improving estimation accuracy of existing methods.



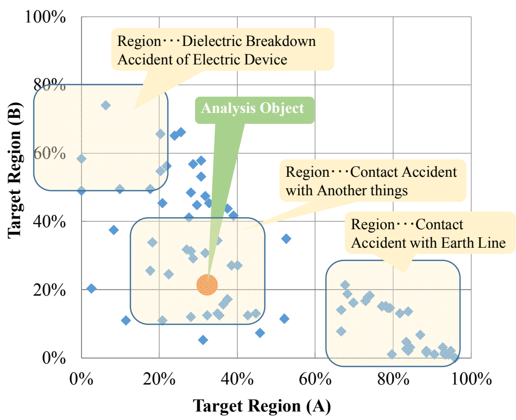
(a) Example of Waveform



(b) Image for Creating Phase Plane Trajectory



(c) Image for Extraction of Waveform Feature



(d) Image for Cause Estimation

Fig.1 Cause Estimation using Phase Plane Trajectory

This paper proposes a method to estimate causes of faults using a statistical analysis that estimates causes of faults without using the collation table.

ESTIMATION FOR CAUSES OF ACCIDENTS

A Statistical analysis to estimate causes of faults is as follows; past waveforms similar to the analysis waveforms are extracted and analyzed the causes of faults with similar past waveforms. Then, possible causes of faults are obtained along with their probabilities. By using this feature, causes of faults can be estimated without creating the collation table of feature quantity and causes of faults.

In the proposed method, cluster analysis is applied. This analysis is a method to divide a data group consisting of n-dimensional sequence of the values into groups by defining the distance between data calculated from the sequence of the values as degree of similarity and combining the similar data. Fig.2 shows the example that past waveform data (12 cases) and waveform data of the analysis target are divided into groups as the result of cluster analysis.

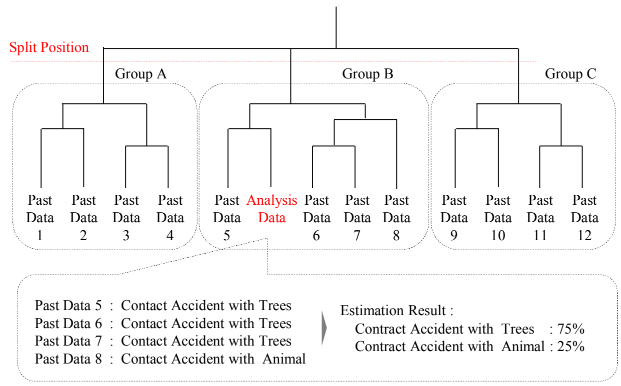


Fig.2. Image for Cause Estimation

Fig. 2 shows that Group B contains four pieces of the past waveform data and one waveform data, which is the analysis target. Assuming that the past waveform data are three cases of tree contacts and one case of animal contact, the cause of the analysis target is likely to be tree contact with a probability of 75 % and animal contact with a probability of 25%.

From the above example, by checking causes of faults, corresponding to the past waveform data in the group the analysis target is included, the cause of the fault is estimated without setting causes of faults for each group in advance.

Fig. 3 shows the steps to estimate causes of faults using statistical analysis.

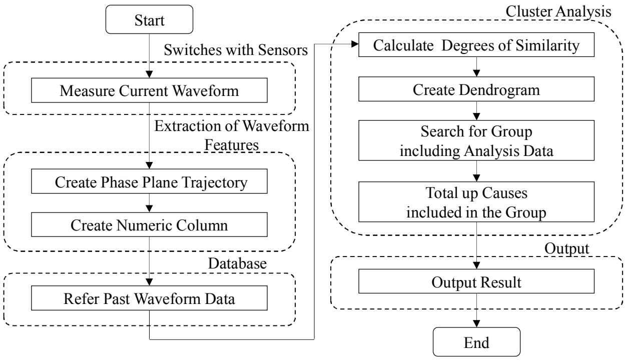


Fig. 3 Procedure of Proposed Method.

In the proposed method to estimate causes of faults, the features of the obtained waveform data are quantified by the phase plane trajectory. In addition, these are divided into groups together with the sequence values of the past waveform data by cluster analysis. Then, the cause of the fault is estimated based on the cause the past waveform data.

VERIFICATION OF PROPOSED METHOD

In the proposed method, the accuracy of estimation may vary depending on the type of cluster analysis applied, so the accuracy of estimation is calculated for each method shown in Table 1 to compare to see how much the accuracy is improved compared to the existing method.

Table 1. Example of Cluster Analysis Methods

Method	How to determine the similarity between the groups.
Nearest Neighbor Method	It compares the data with each other that belong to different groups, and uses the degree of the similarity of the data, which have the highest similarity degree as the degree of the similarity of the group.
Furthest Neighbor Method	It compares the data with each other that belong to different groups, and uses the degree of the similarity of the data, which have the lowest similarity degree as the degree of the similarity of the group.
Group Average Method	It determines the center of gravity of the groups, and uses the degree of the similarity among the center of gravity as the degree of the similarity of the group.
Ward's Method	It uses the amount of change in the degree of the similarity in the group after coupling group, as the degree of the similarity of the group.

For the verification of the proposed method, this paper uses 105 pieces of waveform data with confirmed causes of faults obtained by switches with sensors between May 29, 2012 and December 8, 2015. Of the 105 pieces of waveform data, this paper treats 80 pieces of old data with confirmed causes of faults as the past waveform data, and 25 pieces of new data as newly obtained waveform data to be analyzed. Then, causes of faults are estimated by 25 pieces of waveform data based on 80 pieces of waveform data. Fig.4 shows the distribution of the waveform data, which is used in this study.

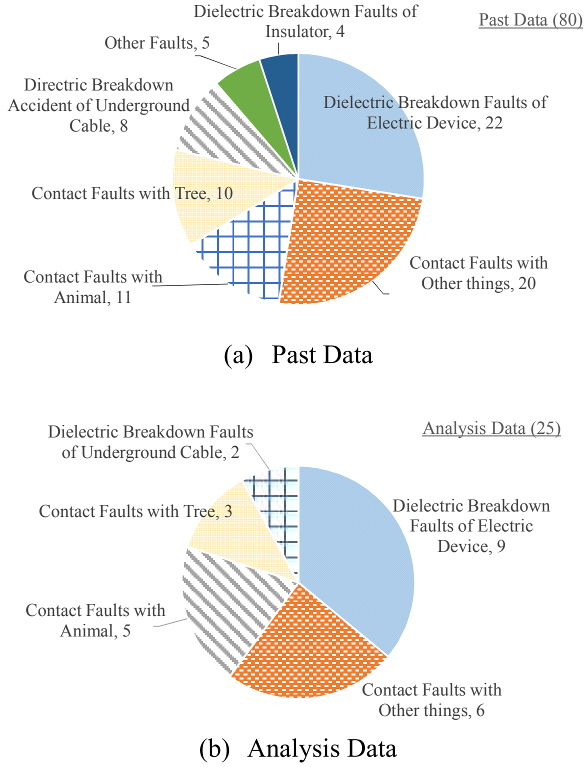


Fig. 4 Data for Verification of Proposed Method

In addition, this paper compares the proposed method with the estimation method by the phase plane trajectory, which is the existing method, but detailed thresholds for estimation of causes of faults are not considered in the past studies, so this proposed method newly establishes the threshold. To be specific, as the example shown in Fig. 5, in the scatter diagram with 25 divided cells ,that are 5 x 5, this method establishes the cause of faults for each cell based on the causes of faults with the past data in the cell.

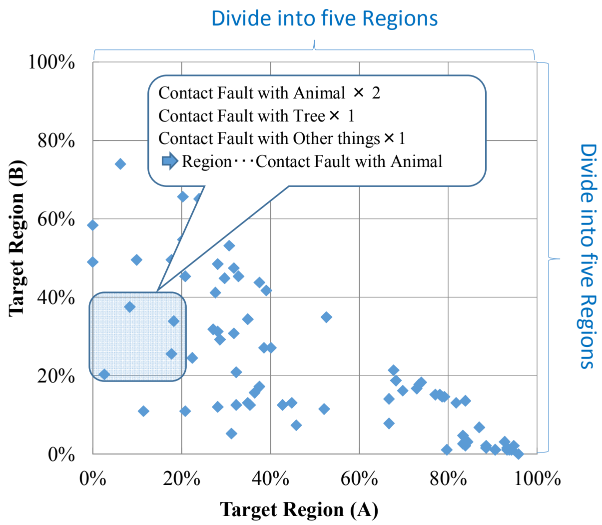


Fig.5 Scatter Diagram of Phase Plane Plots

This paper examines how much the accuracy would improve when cluster analysis is applied to the existing method. Results are shown in Fig.6. In order to compare the accuracy of different cluster analysis methods, next 4 methods are used: the furthest neighbor method, the nearest neighbor method, the Ward's method and the group average method. In addition, the number of groups for classification can also change the accuracy, so the number of partitions are changed between 15 and 60 for the verification of this study.

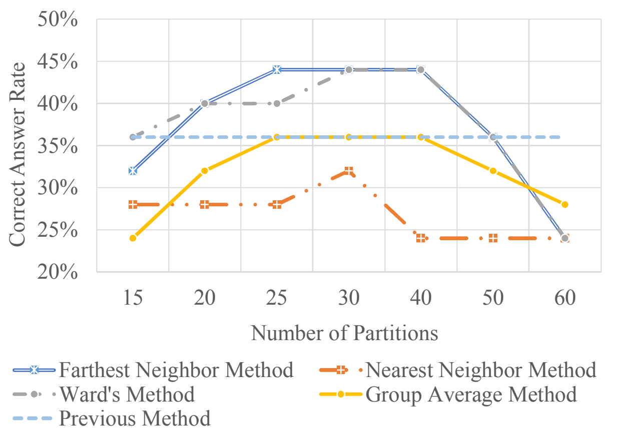


Fig. 6 Comparison of Cluster Methods.

Fig. 6 indicates that the greater the number of partitions is, the better the accuracy becomes in Ward's method and the furthest neighbor method, and the accuracy improvement is expected up to 8% compared to the existing method.

In addition, of the results of cluster analysis of Fig. 6, the example of the furthest neighbor method with the partition number set to 40 is shown in Fig. 7.

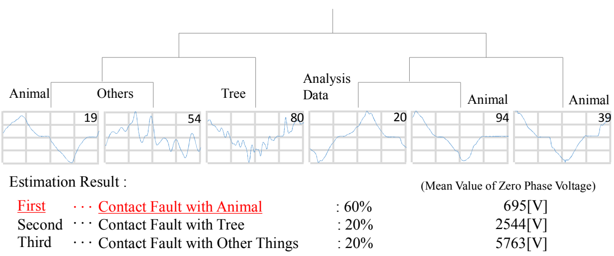


Fig. 7 Example of Cluster Analysis

Fig. 7 is the excerpts of groups containing the analysis targets after cluster analysis is applied. Waveforms similar to the analysis targets are classified near the targets as waveforms 19, 94, 39, that indicates that the waveforms are classified relatively accurately.

On the other hand, the accuracy goes down as the number of partitions becomes great as shown in Fig. 6. It is found that the greater the number of partitions is, the fewer the pieces of data contained in one group becomes, resulting in more cases with no similar past data in the group which contains the analysis target.

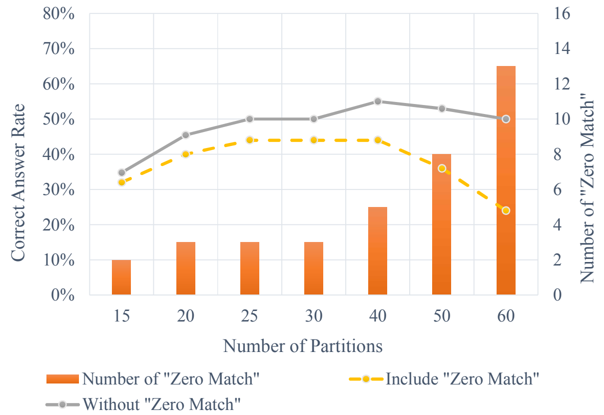


Fig. 8 Relationship of Determination Accuracy and Zero Match

The percentage of correct answers after the exclusion of un-estimated cases is represented in the solid line in Fig. 8. It generally shows an upward tendency as the number of partitions increases.

Next, Fig. 9 indicates the results of estimated causes of faults after excluding the un-estimated cases for different types of faults. For example, in dielectric breakdown fault of electric device as the cause of fault the number of correct answers is six out of the number of waveform data, which is seven.

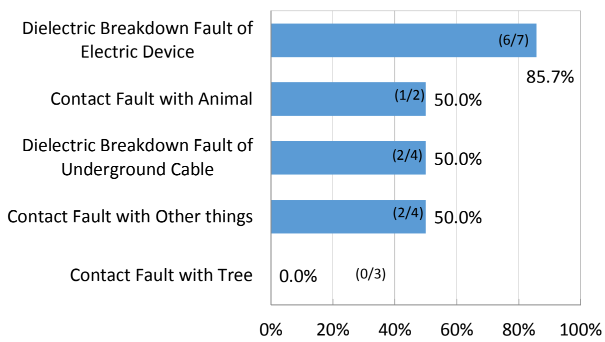
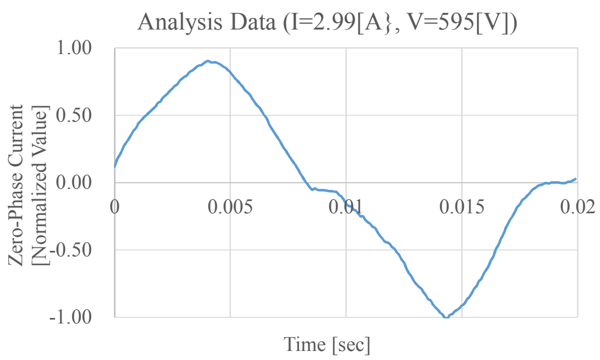


Fig. 9 Estimation Result of Each Causes

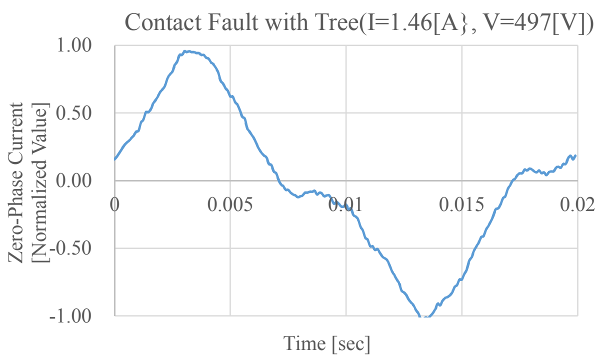
According to Fig. 9, dielectric breakdown of equipment is estimated about 80% accurately, but the percentage is less than 50% for other types of faults. Therefore, the accuracy needs to be improved for practical application.

POUNTS IN IMPROVING ACCURACY

Some results, which turn out to be wrong, indicate that there are some cases, which are difficult to determine the causes of faults by the analysis of waveforms alone. Fig.10 is the case where it is difficult to distinguish the faults by tree contacts or by other things contacts, in which the shapes of the waveforms are visually almost identical and the values of the zero-phase voltage and zero-phase current are close. In order to improve the percentage of correct answers in these cases, it is expected that factors other than waveform data, such as weather condition, geographic information and operation history of relays, need to be included in the analysis.



(a) Analysis Data



(b) Waveforms with Higher Similarities

Fig. 10 Example of Similar Waveforms

CONCLUSION

This paper proposes the method that combines the existing estimation method based on the phase plane trajectory with statistical analysis for estimating causes of faults from current waveform data when a power failure occurs on a distribution line. The accuracy of the proposed method is examined by using 105 waveform data, and the accuracy about 80% is obtained in this study.

On the other hand, for practical application, the estimation results needs to be subdivided and more data need to be accumulated, so continuous studies are necessary. At the same time, in an examination focusing solely on the waveform data, there are cases that is difficult to be subdivided. Therefore, the improvement of the estimation accuracy is expected by combining weather condition and geological information.

With further introduction of switches with sensors to distribution lines, the amount of accumulated data is expected to increase dramatically in the future, so future works continue the research in an effort to improve the estimation accuracy and to subdivide the estimation results.

REFERENCES

[1] M.Okazaki, and T.Inaba, 1992, “Fault Locating Method on Used Harmonic Components of Arc Voltage Wave Form ", *T.IEE Japan*, Vol.112-A, No.9, 757-762.

[2] T.Horita, M.Sumiyoshi, T.Wakai, N.Ikeda, and I.Kitamura, 1997, "A New Classification Method of Ground Faults in Distribution Lines with Phase Plane Trajectory of Their Waveforms", *T.IEE Japan,* vol. 117-B, 196-202.

[3] M.Watanabe, K.Kotanshi, O.Nakamura, H.Kurioka, S.Hukui, and K.Tuji, 1995, "A New Method for Discrimination of Ground Faults Causes in Distriburion Lines", *T.IEE Japan*, vol. 115-B, No.1, 18-23.

[4] W. Zhao, Y. H. Song, and Y. Min, 2000, “Waveform characteristics of underground cable failures”, Elsevier Science Journal of Electric Power Systems Research, Vol.53, No.1 pp.23-30.

[5] S. Kulkarni, D. Lee, A. J. Allen, S. Santoso and T. A. Short, 2010, “Waveform characterization of animal contact, tree contact, and lightning induced faults”, IEEE PES General Meeting, pp.1-7.