Research on Adaptive Power Transmission Line Fault Inspection

Jian Wu, Shuang Shao, Dongping Jiang, Hongfang Yao, Caiguo Ma, Kekui Xu*
Hangzhou power supply company of State Grid Zhejiang Electric Power Co., Ltd, Hangzhou City of Zhejiang Province, 310000
xkk@hongsong.cn

Abstract. Power transmission line is one of the most important infrastructures of power system, and its safety monitoring is of great significance. The conventional way of fault monitoring of power transmission lines by only setting threshold value on single temperature data of strain clamp turned out to multiple misjudgements and delayed alarms, causing the increment of operation risk of power transmission line. In this paper, various types of time-varying sensors data such as strain clamp temperature data, environmental data, and cable ampacity data are accounted. Also, an unsupervised machine learning algorithm - K-means clustering algorithm was introduced to build a discriminant model in detecting the defects of power transmission line. Experimental results proved that the proposed method is able to avoid delayed alarms as well as misjudgement incurred from conventional method. As a result, the operation safety of power transmission line and inspection efficiency can be improved. The inspection cost would be reduced as well.

1. Introduction
At the background of “Three Types and Two Networks” from State Grid Corporation of China, operation safety of power transmission plays a very important role in power grid system [1-2]. As an important infrastructure of power system, power transmission line has long mileage and wide coverage. The integrity and safety of its structure is related to all aspects of production and life. Therefore, it is one of the important engineering structures, and has the important monitoring necessity.

Major defective influential includes increment of load current, worn and oxidized line, broken core as well as impact from foreign objects [3-4]. During power transmission, when power transmission line operates within rated load, temperature of line core is fine to within a certain range; if transmission line is overloaded, line core’s temperature will increase rapidly and may lead to system breakdown and cause unwanted accident. On the other hand, worn burrs and broken core also cause line to increase in temperature. Therefore, monitoring temperature of transmission line is very significant in failure monitoring of transmission line in condition such as increase of current load, worn or oxidized line and broken core [5-7].

When the temperature of transmission line increases, some heat is transferred to strain clamp and causes strain clamp to increase in temperature. Conventional method accounted only a single temperature feature, setting a threshold value on strain clamp temperature as failure monitoring. Yet, line cores’ temperature would not increase and breaches threshold value in a short period, but has an increasing trend at the beginning of power transmission line failure. Meanwhile, conventional method on failure monitoring unable to automatically diagnose the condition of the power transmission line.
from data trend, which means manpower is needed to run the analysis which leads to decrement of efficiency. In addition, conventional method also causes mis-reported and delayed alarms [8-10].

The method proposed in this paper uses environmental data, cable ampacity data and so on that putting abrupt abnormalities into consideration, taking multiple sensors data such as power transmission line strain clamp temperature, environmental data and cable ampacity data with the adoption of K-mean clustering algorithm and clusters’ condensation to build a discriminant model [11-12] acting as a warning to power transmission line defective call. This model can avoid misjudgment and delayed alarm of current techniques, improve the operation safety of power transmission line, also at the same time, the cost of manpower on inspection would be reduced and improve work efficiency.

2. Fault Monitoring by K-means Clustering
K-means clustering algorithm is an unsupervised learning method which manipulates multiple iterative processes in order to distinct data into predefined, \( K \) groups or clusters. Data is predefined into \( K \) groups, and then randomly choose \( K \) samples as centroids of clusters; compute the distance between sample and centroid is minimum. The application of K-means clustering algorithm is simple yet effective, parameter to fine-tune is the number of clusters, \( K \).

Various sensors data such as power transmission line strain clamp temperature, environmental data and cable ampacity data is made as samples, which formed training set \( A \). Training set \( A \) need to go through data preprocessing step to remove duplicates and anomaly data; then apply K-means clustering algorithm on processed training set \( A \), extract the important features from dataset on defective power transmission line, calculate the best cluster number \( K \). Each data in training set is distributed into multiple clusters, and this creates the monitoring discriminant model.

The optimum value \( K \) is resulted after building discriminant model, for each cluster, the distance between sample and centroid can be calculated, as well as the sum of distance. Trend of the calculated distance between cluster centroid and sample can be plotted and discriminant threshold can be captured from the plot.

2.1. Choosing the Best Cluster Number, \( K \)
Silhouette Coefficient is an important metrics in clustering task to measure cohesion in within and separation between clusters. Sample points in within the cluster need to be closed together while keeping distance between other clusters. For training set \( A \), we applied Silhouette Coefficient formula to calculate the optimum cluster number, \( K \), as shown in (1).

\[
s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}
\]  

(1)

\( a \) is the mean distance of intra-cluster, average dissimilarity of other sample points in within the same cluster; on the other hand, \( b \) is mean distance to the closest cluster other than its cluster. According to the equation above, range of \( s \) is between -1 to 1. When \( s \) is close to -1, meaning that \( s \) should be drawn to other cluster or vice versa; larger value \( s \) showing better intra-cluster cohesion and inter-cluster separation.

2.2. Calculate Discriminant Threshold
After the clustering computation, valid sample points are then appointed to \( K \) clusters. Distance between all valid samples towards their corresponding cluster centroid can be calculated, then sum and sort; converts the computed result into a visual graph. The training set \( A \)’s sample count is set as x-axis while the distance between cluster centroid as y-axis. The rapid change of trend or an obvious change of gradient / slope is made inflection point of training set and the value of y-axis corresponds to the inflection point is the discriminant threshold, \( y \) for unknown dataset discrimination [13-14].
3. Building of Discriminant Model

3.1. Data Features
Various sensors data is needed in order to build the discriminant model, e.g. power transmission lines’ strain clamp, environmental data, cable ampacity data, etc.

(1) Strain clamp temperature data is a real-time data which collected from temperature sensor located at power transmission line tower pole; the strain clamp temperature of defective line routes would have explicit features.

(2) Environmental data is real-time data collected from ambient monitor sensor at surroundings of power transmission line. Worse surrounding would cause defects of power transmission line. For example, high wind speed (scratched risk), high sun intensity (high temperature on power transmission line), too high or too low (frozen) surrounding temperature would cause defective failure of power transmission line.

(3) Cable ampacity data is real-time data collected from current sensor on power transmission line.

Table 1. Features of Power Transmission Line

| Data Type        | Details                                                                 |
|------------------|-------------------------------------------------------------------------|
| Strain Clamp Temperature Data | Daily average strain clamp temperature                                   |
|                  | Daily standard deviation strain clamp temperature                        |
|                  | Daily average high load strain clamp temperature                          |
|                  | Daily standard deviation high load strain clamp temperature              |
|                  | Daily median high load exceeded strain clamp temperature’s period        |
|                  | Daily average high load exceeded strain clamp temperature’s period       |
|                  | Daily average low load strain clamp temperature                          |
|                  | Daily standard deviation low load strain clamp temperature               |
|                  | Daily median low load exceeded strain clamp temperature’s period         |
|                  | Daily average low load exceeded strain clamp temperature’s period        |
|                  | Daily standard deviation low load exceeded strain clamp temperature’s period |
| Environmental Data | Ambient wind speed                                                      |
|                  | Intensity of the sun                                                     |
|                  | Ambient temperature                                                      |
| Cable ampacity data | Transmission line conductor current                                     |

3.2. Model Fitting
K-means clustering algorithm is adopted in our experiment; using the historical data (training set A) of various sensors such as power transmission lines’ strain clamp temperature data, environmental data, cable ampacity data and so on as training sample set; performs data cleaning, model fitting and evaluation as well as threshold determination steps to build the discriminant model, as illustrated in Fig. 1.
4. Experiment Validation

4.1. Experiment Settings and Data Processing

Based on Code for design of 110kV~750kV overhead transmission line (GB 50545-2010) and conductors of insulated cables (GB/T3956-2008), a 220kV voltage is added on steel strain transmission line to simulate low and high temperature conditions; specification of the steel strain transmission line use in experiments: length of 1km and diameter of 120mm². Four different scenarios were put into test to evaluate the accuracy of discriminant model.

The State Grid Zhejiang Hangzhou Yuhang District Power Supply Company provided a thousand historical data of power transmission lines’ strain clamp temperature, environmental data, cable ampacity data and various sensors data. The sample set first remove sensitive fields, and then performs computation as shown in Figure 1. Figure 2 illustrated the changes of strain clamp temperature and cable ampacity data at different load. The first 300 data corresponding to regular data load; mid 400 data corresponding to high data load; while the last 300 data corresponding to low data load. Environmental data is shown in Fig. 3, which include ambient temperature, sun intensity and wind speed. First 500 data correspond to cold environment while last 500 data correspond to high temperature environment.

![Figure 2. Strain clamp temperature and cable ampacity data.](image1)

![Figure 3. Environmental data.](image2)
4.1.1. Data Cleaning. From the average dissimilarity dimension 1 and minimum dissimilarity dimension 2 in within a same cluster, by adopting K-means clustering algorithm on historical data of each data point in training set, we can get the optimum cluster number, $K = 3$ and its centroid. In addition, the distance between each sample points and the corresponding the clusters’ centroids can be calculated. If the Euclidean distance is too large, meaning that sample point is away from its cluster centroid which can be annotated as outlier / anomaly samples.

There is 126 valid data points after removal of anomaly samples, shown in Table 2. 9 out of 126 data point is data collected at abnormal condition.

Table 2. Four different specification of steel transmission line

| Line Code | Specification                                                                 |
|-----------|-------------------------------------------------------------------------------|
| 1         | Regular 1km multi-core steel transmission line                                 |
| 2         | 1km multi-core steel defective transmission line (grinded apparent burrs using sandpaper) |
| 3         | 1km multi-core steel defective transmission line (immersed in acid rain-like solution, with sign of oxidation on surface) |
| 4         | 1km multi-core steel defective transmission line (clamped two cores)           |

4.1.2. Model Analysis. The plot of various sensors value versus feature importance is shown in Figure 5. The difference between 12 temperature features in the three clusters is clearly illustrated. To further analyses, we can capture some characteristics from the six features which are the daily average of high load strain clamp temperature, daily standard deviation of high load strain clamp temperature, daily average of low load strain clamp temperature, daily standard deviation of low load strain clamp temperature, daily average of strain clamp temperature and daily standard deviation of strain clamp temperature. Overall trend of Cluster 1 and Cluster 2 is similar to each other; also, feature value Cluster 2 is relatively smaller than Cluster 1 but Cluster 3 has an opposite trend compared to trend of Cluster 1.

![Figure 4. Clusters distribution](image)

Table 3: Distribution of valid data points

| Cat.   | Cluster 1 | Cluster 2 | Cluster 3 | Total |
|--------|-----------|-----------|-----------|-------|
| Sample count | 25        | 64        | 37        | 126   |
Figure 5. Feature values of three clusters

For environmental data, feature value corresponds to Cluster 1 is relatively low, biased towards low temperature environment; Cluster 2 and Cluster 3 are quite close to each other but feature value of Cluster 3 is slightly higher, biased towards high temperature environment.

The strain clamp temperature in Cluster 2 has severe defect characteristics; while in Cluster 1 and Cluster 3, the characteristics are similar to common defect of power transmission line, as illustrated in Fig. 3. Conventional way of using only strain clamp temperature to diagnose defect is not quite accurate.

| Data Type                          | Cluster 1                                      | Cluster 2                                      | Cluster 3                                      |
|------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| High load – Daily average strain   | Normal, but period of high load is short      | Highest, serious exceeding at high load        | Slight high, exceeded normal temperature       |
| average strain clamp temperature   |                                               | condition                                     |                                               |
| Low load – Daily average strain    | High, temperature is slightly high at low     | Slight high, higher than normal condition      | Normal, temperature does not change much      |
| average strain clamp temperature   | load condition                                |                                               |                                               |
| Daily average strain clamp         | Huge change, line overall indicators are      | Higher, overall indicators are higher          | Slight high, similar to common defect condition|
| temperature                        | normal                                        |                                               |                                               |
| Environmental Data                | Slight low, biased towards cold environment   | Normal                                        | Slight high, biased towards high temperature   |
| Conclusion                         | period of high load is short, abnormal        | Overall temperature is high, showing the sign of severe defect |
|                                    | temperature value                             |                                               | Common defects                                 |
4.1.3. **Threshold.** After the centroid of each clusters is finalized, compute the sum of distance between sample points \( x \) and centroid; the computation result is sorted and plot into graph as illustrated in Fig. 6.

![Figure 6. Plot of Incremental Distance](image)

In Figure 6, x-axis is the number of sample data and y-axis is sum of distance between sample data and centroid. When \( x \) is equal or lesser than 117, the speed of distance increment is gentle; while \( x \) is greater than 117, the increment slope become very steep. From the plot of Figure 6 we can conclude, \( x = 117 \) is the inflection point of training dataset and the corresponding distance value of 2.25 is the threshold for our method.

When sum of distance between a sample data point to the three cluster centroids is equal or greater than the threshold value, then we can determine the power transmission line has defect risk or vice versa.

4.2. **Low Temperature Environment’s Experiment Evaluation**

The experiment is run at condition of ambient temperature of 0°C, sun intensity 40%-60%, wind speed 1.6-3.3m/s, voltage 220kV, unit of experiment resistant is Ω/km, as shown in Table 5.

| Line Code | Resistance | Conventional Method Temperature | New Method’s Feature Value | Line Code | Resistance | Conventional Method Temperature | New Method’s Feature Value |
|-----------|------------|--------------------------------|---------------------------|-----------|------------|--------------------------------|---------------------------|
| Initial Value 4th hour | | | | 4th hour | | | |
| 1 | 0.255 | 2.31 | 0.67 | 1 | 0.256 | 3.75 | 0.69 |
| 2 | 0.268 | 2.42 | 1.87 | 2 | 0.599 | 10.5 | - |
| 3 | 0.283 | 2.43 | 2.13 | 3 | 0.601 | 11.7 | - |
| 4 | 0.371 | 2.33 | 2.27 | 4 | 0.371 | 3.65 | - |
| 2nd hour | | | | 6th hour | | | |
| 1 | 0.255 | 3.12 | 0.68 | 1 | 0.258 | 3.90 | 0.67 |
| 2 | 0.327 | 4.23 | 2.44 | 2 | - | - | - |
| 3 | 0.316 | 3.97 | 2.39 | 3 | - | - | - |
| 4 | 0.371 | 3.56 | - | 4 | 0.371 | 3.67 | - |

According to current norm and standards, conventional method sets the warning threshold when strain clamp temperature increases at equal or greater than 10 while our method taking the sum of distance between predicting sample and centroids at equal or greater than 2.25 as warming threshold, we can conclude:

(1) According to Table 5, Line 1 is regular transmission line, it does not turn on any warning from two methods.
(2) According to Table 5, for Line 2 and Line 3, our method already turned-on warning during the 2nd hour but conventional method only output warning at 4th hour which the resistance is twice the normal resistance value.

(3) According to Table 5, Line 4 gets no warning by conventional method but our method called for warning after startup.

4.3. High Temperature Environment’s Experiment Evaluation

The experiment is run at condition of ambient temperature of 50°C, sun intensity 40%-60%, wind speed 1.6-3.3m/s, voltage 220kV, unit of experiment resistant is Ω/km, as shown in Table 6.

| Line Code | Resistance | Conventional Method Temperature | New Method’s Feature Value | Line Code | Resistance | Conventional Method Temperature | New Method’s Feature Value |
|-----------|------------|---------------------------------|---------------------------|-----------|------------|---------------------------------|---------------------------|
| 1         | 0.263      | 39.1                            | 0.92                      | 1         | 0.271      | 51                             | 0.96                      |
| 2         | 0.279      | 39.0                            | 1.87                      | 2         | -          | -                              | -                         |
| 3         | 0.288      | 38.8                            | 2.13                      | 3         | -          | -                              | -                         |
| 4         | 0.390      | 39.2                            | 2.44                      | 4         | 0.413      | 52.3                           | -                         |

| Line Code | Resistance | Conventional Method Temperature | New Method’s Feature Value | Line Code | Resistance | Conventional Method Temperature | New Method’s Feature Value |
|-----------|------------|---------------------------------|---------------------------|-----------|------------|---------------------------------|---------------------------|
| 2nd hour  |            |                                 |                            | 6th hour  |            |                                 |                           |
| 1         | 0.263      | 45                              | 0.93                      | 1         | 0.265      | -                              | 0.95                      |
| 2         | 0.356      | 50.3                            | 2.94                      | 2         | 0.599      | 10.5                           | -                         |
| 3         | 0.382      | 52.2                            | 2.87                      | 3         | 0.601      | 11.7                           | -                         |
| 4         | 0.395      | 44.2                            | -                         | 4         | 0.265      | -                              | 0.95                      |

(1) According to Table 6, Line 1 is regular transmission line case, conventional method passing false alarm at 4th hour.

(2) According to Table 6, for Line 2 and Line 3, our method already startup warning during the 2nd hour but conventional method only output warning at 6th hour which the resistance is twice the normal resistance value.

(3) According to Table 6, Line 4 gets pass warning after 4th hour according to increase of ambient temperature by conventional method but our method called for warning right after startup.

5. Conclusion

As a conclusion, the fault monitoring method by K-means clustering algorithm and cluster condensation performed better than conventional method, grinded and oxidized conditions (Line 2 and 3) are captured at the 2nd hour in both low and high temperature environment. As contrast, conventional method only output warning at the 4th hour and the 6th hour in both environments respectively. Our method called off warning right after startup in short of cores case (Line 4) but conventional method only returned warning at the 4th hour in high temperature condition while no defect detected in low temperature condition. The results show that new method introduced in this paper successfully avoided mis-judgement and delayed alarm from conventional method as well as improved safety of transmission line operation.

References

[1] J Wang, X Xiong, J Hua and X Lu Y 2019 International Journal of Electrical Power & Energy Systems Safety strategy of power transmission channel coordinated with transfer capability support for power system emergency 110 232-245

[2] A. Kudzys Y 2006 Engineering Structures Safety of power transmission line structures under wind and ice storms 28(5) 682-689
[3] C Sun, X Wang and Y Zheng Y 2019 *International Journal of Electrical Power & Energy Systems* Data-driven approach for spatiotemporal distribution prediction of fault events in power transmission systems 113 726-738

[4] M Parsi, P Crossley, P L Dragotti and D Cole *Electric Power Systems Research* Wavelet based fault location on power transmission lines using real-world travelling wave data 186 10626

[5] Y Hao, Y Cao, Q Ye, H Cai and R Qu Y 2015 *Optik - International Journal for Light and Electron Optics* On-line temperature monitoring in power transmission lines based on Brillouin optical time domain reflectometry 126(19) 2180-3

[6] M Kunicki, S Borucki, D Zmarzły and J Frymus *Measurement* Data acquisition system for on-line temperature monitoring in power transformers 161 107909

[7] R Lecuna, P Castro, M Manana, A Laso, R Domingo, A Arroyo and R Martinez Y 2020 *Electric Power Systems Research* Non-contact temperature measurement method for dynamic rating of overhead power lines 185 106392

[8] G Sun, X Li, J Wu, R Chen and G Chen Y 2020 *Engineering Structures* Deformation of stainless-steel cables at elevated temperature 211 110498

[9] T Jiao, Z Zhou, J Liu, H Xiao and J Ou Y 2020 *Measurement* Large strain-tolerated smart steel strand with built in coaxial cable Fabry–Perot interferometer 151 107019

[10] F Han, D Dan, Y Zou and H Lei Y 2020 *Mechanical Systems and Signal Processing* Experimental and theoretical study on cable-supporting system 140 106638

[11] S Tayfur, N Alver, S Abdi, S Saatci and A Ghiami Y 2018 *Engineering Fracture Mechanics* Characterization of concrete matrix/steel fiber de-bonding in an SFRC beam: Principal component analysis and k-mean algorithm for clustering AE data 194 73-85

[12] A Ahmad, and L Dey Y 2007 *Data & Knowledge Engineering* A k-mean clustering algorithm for mixed numeric and categorical data 63(2) 503-27

[13] J Żygierewicz, C Siciużycki, R König and P J Durka Y 2008 *Journal of Neuroscience Methods* Event-related desynchronization and synchronization in MEG: Framework for analysis and illustrative datasets related to discrimination of frequency-modulated tones 168(1) 239-247

[14] Z Pan, Y Wang and Y Pan Y 2020 *Knowledge-Based Systems* A new locally adaptive-nearest neighbor algorithm based on discrimination class 204 106185