

# K-NEAREST NEIGHBORS ALGORITHM APPLICATION IN THE ELECTRICAL GRID STATES RECOGNITION PROBLEMS

A.L. Kulikov<sup>1</sup>, D.I. Bezdushniy<sup>1,\*</sup>, V.Yu. Osokin<sup>1</sup> and A.A. Sevostyanov<sup>1</sup>

<sup>1</sup>Nizhniy Novgorod State Technical University, Russia

**Abstract.** One of the ways to improve the reliability of electrical grids is associated with the introduction of complex and resource-intensive algorithms in intelligent electronic devices (IED) that perform the functions of relay protection and automation at substations. Simulation modelling is used to study the features of the protected object functioning and its application makes it possible to take into account the variability of electrical grid states in the formation of IED algorithms, which are characteristic of the analyzed electrical grid section. In addition, this approach makes it possible to use for short circuits detection only those information features that have a high information value in a specific problem of states recognition. Machine learning methods are advisable to use for modern relay protection algorithms implementation. One of such methods is the k-nearest neighbours method. The article substantiates the effectiveness of the method application in comparison with the conventional algorithms on the example of protection of an electrical grid section with a distributed generation source. The reported study was funded by RFBR, project number 19-38-90144.

## 1 Introduction

Recognition and localization of short circuits (SC) in electrical grids by relay protection devices are essential components of trouble-free operation of power systems. The current trend towards digitalization of secondary substation systems, including relay protection and automation, makes it possible to implement much more complex and potentially more efficient algorithms for recognizing emergency states in comparison with existing technical solutions.

A promising direction in relay protection is the informational approach [1, 2]. It involves the use of the states simulation modeling results to form protection algorithms. An important feature of the information approach is the selection of the features controlled by the protections individually for each separate problem of states recognition, based on the impurity function. As a rule [1, 2], the features are combined into a system of setting planes with operating zone formed by the simulation modeling and training results.

However, the application of setting planes is not the only way to form a decision rule for relay protection using simulation modeling. The application of special machine learning algorithms can be proposed as an alternative approach. Among the problems solved by machine learning tools, there is the classification problem. It lies in assigning some feature vector  $\mathbf{x}$  to one of the given classes  $Y_1 \dots Y_m$  based on a training sample. This sample consists of a set of vectors  $\mathbf{x}_1 \dots \mathbf{x}_n$ , which class is known. As can be noted, the classification problem is identical to the informational approach problem in relay

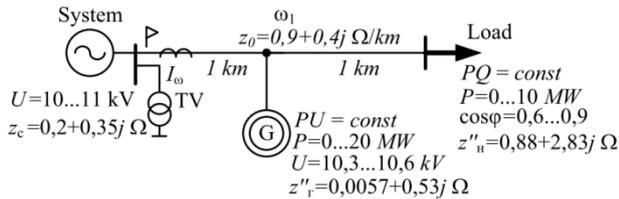
protection. In this case, various electrical state parameters act as features and they are available for measurement at the substation. Many classes in turn are formed by the controlled states of the protected object (normal state, self-starting state, short circuit state, etc.), and the simulation modeling results become the training sample. Thus, the machine learning is used to identify emergency states of power facilities functioning and makes it possible to develop new algorithms of relay protection. Moreover, information features are combined into a single feature space of arbitrary dimension. This contributes thereby to more efficient use of available information about the current electrical grid state. In this regard, it is important to study new ways of relay protection implementation using machine learning algorithms, as well as to assess their recognition ability.

Let us analyze the k-nearest neighbors method application, which implements the classification based on the estimation of the distance function between the classified object and the objects of the training sample. A similar approach, which also involves the distance function calculation, was previously used in relay protection problems [3] to determine the recognizability of the observed states in a multidimensional feature space when implementing the functions of long-range redundancy of transformer relay protection.

Let us develop a relay protection algorithm for an electrical grid section with a distributed generation source G (Fig. 1) using the k-nearest neighbors method. Consider a separate function of the relay protection device, which provides protection of the  $\omega_1$  line from three-phase and phase-to-phase short circuits. At the

\* Corresponding author: [dimok.93@mail.ru](mailto:dimok.93@mail.ru)

same time, its operation is excluded in operating states, as well as in self-starting states of the load.



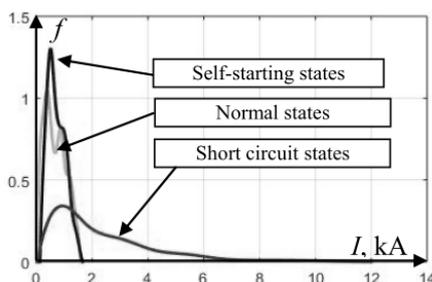
**Fig. 1.** Scheme of the protected electrical grid section

## 2 The procedure for simulation experiments implementation

The purpose of simulation modeling is to form a training sample for the classification algorithm that reliably reflects the set of potentially possible controlled states. Therefore, the simulation model should take into account the inconstancy of some parameters of the equivalent circuit, such as voltage and resistance of the grid, load power, generation power, etc. When modeling a short circuit state, the variable values are also the distance from the fault point and the value of the transient resistance. When carrying out simulation modeling, it is advisable to use the Monte Carlo method [4]. It consists in repeated reproduction of experiments in such a way that at each iteration all variable parameters take random values within the ranges specified during model tuning. Figure 1 shows both constant and variable parameters of the analyzed electrical grid section.

Let us simulate the operating states of the considered circuit, as well as the self-starting and short circuit states on the  $\omega_1$  line. As a result, we obtain complexes of branch currents and node voltages of each of the phases in each of the states. The data array generated in this way makes it possible to analyze the effectiveness of various types of relay protection devices to ensure protection of the  $\omega_1$  line.

For example, to analyze the recognition ability of current protection, we will implement statistical distributions of the effective value of the current of one of the phases in all modeled states (Fig. 2).



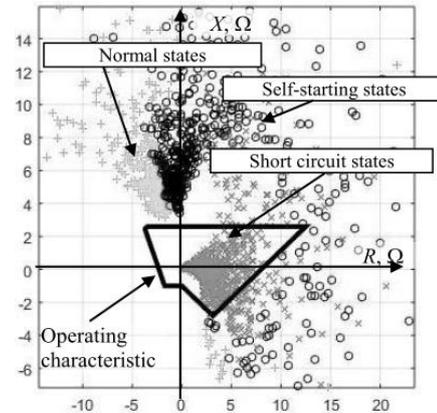
**Fig. 2.** Phase current distribution in controlled states

As can be seen from Figure 2, a significant part of the fault currents turned out to be comparable in magnitude with the currents of normal states and self-starting load states. This circumstance is due to the fact that the generator, included in the branch line, can reduce the

fault current flowing through the installation point of the protection and so decrease its sensitivity. It is obvious that the use of current protection in the analyzed grid section is ineffective.

Let us study the effectiveness of the distance protection application for the relay protection organization of  $\omega_1$  line (Fig. 1).

Taking into consideration the previously obtained simulation modeling results, we shall calculate the complex resistances values estimated at the installation point of the protection for each of the experiments, and place them on the complex plane (Fig. 3).



**Fig. 3.** Complex impedance measurements in the analyzed states

Distance protection characteristic (Fig. 3), obtained by the condition of a complete prohibition on operation in normal and self-starting load states, is capable of disconnecting the fault on the line with a probability of 74%. Despite the fact that the probability of fault recognition by distance protection significantly exceeds the current protection probability, its effectiveness is still insufficient for reliable emergency states detection of an electrical grid section (Fig.1). A further increase in the recognition ability of protection can be achieved as a result of using machine learning algorithms, in particular, the k-nearest neighbors method.

## 3 Principles of the k-nearest neighbors algorithm application

Among the many well-known machine learning algorithms, the k-nearest neighbors method has the most intuitive operating principle. According to the method [5], the classification of an object (state) is carried out in accordance with the most common class among its "neighbors", that is, the training sample objects (states) located at a minimum distance from the classified object. The number of analyzed neighbors is random and it is defined based on the requirements of a specific classification problem when choosing the parameters of the simulation model. Various distance functions can be used when implementing the method. The following metrics are most often used:

Euclidean metric: 
$$D(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2} \quad (1)$$

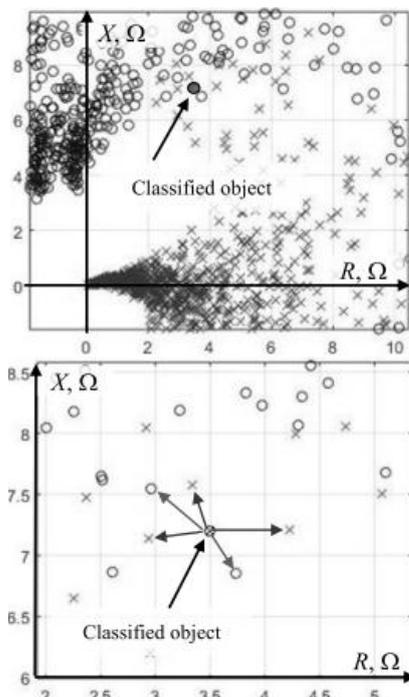
Chebyshev distance: 
$$D(\mathbf{x}, \mathbf{y}) = \max_{i=1..m} |x_i - y_i| \quad (2)$$

Manhattan distance: 
$$D(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^m |x_i - y_i|, \quad (3)$$

where  $\mathbf{x}=[x_1, x_2...x_m]$ ,  $\mathbf{y}=[y_1, y_2...y_m]$  are vectors in m-dimensional space, the distance between these vectors needs to be determined;  $D(\mathbf{x}, \mathbf{y})$  is the required metric.

Let us research the classification algorithm for the relay protection device by the k-nearest neighbors method in the feature space formed by the resistance and reactance (Fig. 4). The choice of the k value is usually done empirically. Let us take this value as a first approximation equal to 5. As a distance function we choose the Euclidean metric. Let us denote the SC states on  $\omega_1$  as  $\alpha$ -states and combine the normal states with the self-starting load states and denote the obtained set as  $\beta$ -states.

Let us find  $k = 5$  values in the training sample and define their classes. The distance from such values to the classified impedance value (expression (1)) is minimal. According to fig.4, among the nearest neighboring values of the impedance, there were 3 objects applying to the class  $\alpha$  and 2 objects characteristic of the class  $\beta$ . Thus, in accordance with the majority principle, the current analyzed state apply to the class  $\alpha$ .

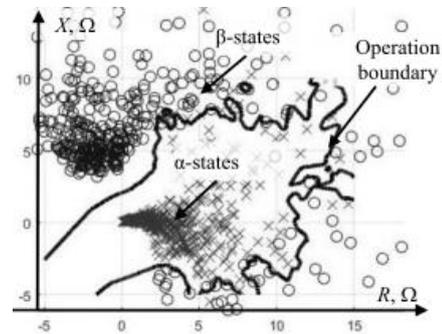


**Fig. 4.** The impedance sample value classification in the feature space

Let us select on the setting plane (in the feature space) operation region of the protection, inside which the state

will belong to the class  $\alpha$  when implementing the researched algorithm (Figure 5).

Also let us estimate the efficiency of the relay protection recognition algorithm using the error matrix [6] containing the correct classification probabilities of each of the classes, as well as the type I and type II errors probabilities. We shall randomly divide the aggregate set of model simulation experiments into two groups, one of which will be used to train the protection algorithm, and the second - for its subsequent testing.



**Fig. 5** The operation region for k-nearest neighbors algorithm.

The error matrix as applied to the conditions of the k-nearest neighbor algorithm study corresponds to Table 1.

**Table 1.** Error matrix corresponding to the k-nearest neighbors algorithm

		Predicted class	
		$\alpha$	$\beta$
True class	$\alpha$	95,4 %	4,6%
	$\beta$	2,8%	97,2%

The analysis of Table 1 shows that the k-nearest neighbors algorithm application provides the operation of the relay protection in 95.4% of the faults on the protected line. Moreover, a false operation is possible with a probability of 2.8%.

#### 4 The k-nearest neighbors algorithm modification

The majority principle used in the k-nearest neighbors algorithm assumes that each of the training sample classes equally affects the decision making in the recognition process. As a result, the number of type I and type II errors [7] allowed by the classification algorithm is approximately the same. This principle is often not acceptable when relay protection organization. The protection devices are parameterized in such a way as to completely exclude the possibility of non-selective operation, even if the sensitivity of the relay protection reduces. Consequently, it assumes that the cost of the type I error is significantly higher than the type II one. Thus, when implementing the k-nearest neighbors algorithm, it is advisable to change the rule for an object classification, depending on the classes of its neighbors. Instead of the majority principle, we shall accept the following classification rule (criterion): an object belongs to the class  $\alpha$  only if each of its k nearest

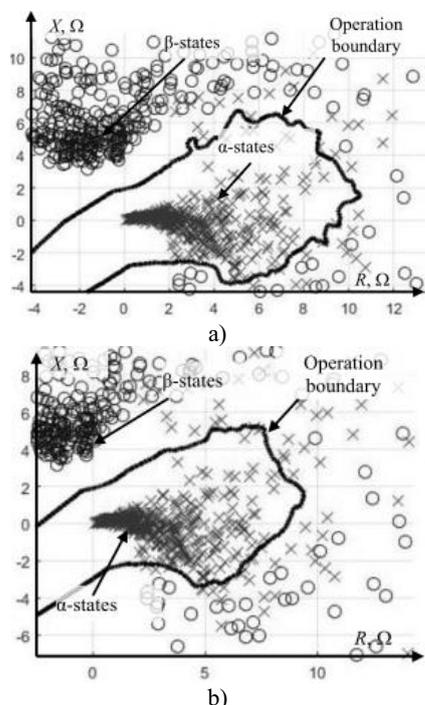
neighbors also belongs to the class  $\alpha$ . Otherwise, the object belongs to class  $\beta$ .

The accepted criterion will make it possible to exclude the likelihood of a false assignment of the object functioning state to a set of  $\alpha$ -states, thereby provoking a non-selective operation of the relay protection device. The operation regions obtained as a result of the implementation of the modified k-nearest neighbors algorithm to the previously obtained training sample are shown in Figure 6 (a) and the components of the corresponding error matrix are shown in Table 2.

**Table 2.** Error matrix corresponding to the modified k-nearest neighbors algorithm (k=5)

		Predicted class	
		$\alpha$	$\beta$
True class	$\alpha$	92,8 %	7,2%
	$\beta$	0,4%	99,6%

As can be seen (Table 2), due to the decision rule modification at the cost of some loss of sensitivity, it was possible to reduce the false operation probability of the protection algorithm from 2.8% to 0.4%.



**Fig. 6.** Operation region for the modified k-nearest neighbors method; a) k=5; b) k=15

It is possible to reduce the number of false operations of the protection in operating and self-starting load states by increase in the k value. Then, the decision shall be based on a larger number of the training sample values, which, in turn, shall shift the boundary of the operating characteristic further from the points of the complex plane typical for  $\beta$ -modes.

Figure 6 (b) illustrates the operation region of the protection algorithm using the modified k-nearest neighbors algorithm and k = 15. In this case, the error matrix takes the form corresponding to Table 3:

**Table 3.** Error matrix corresponding to the modified k-nearest neighbors algorithm (k=5)

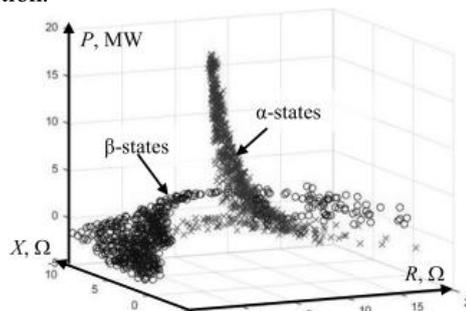
		Predicted class	
		$\alpha$	$\beta$
True class	$\alpha$	88,6 %	11,4%
	$\beta$	0%	100%

An increase in the k value (Table 4) led to a decrease in the protection sensitivity to 88.6%, as well as a decrease in the probability of false operation to zero.

### 5 Change in the composition of information features

The protection algorithm, implemented on the basis of the k-nearest neighbors method, uses the resistance and the reactance as information features and makes it possible to recognize 88.6% of faults on the line, which is 14.6% more than distance relay protection using similar input magnitudes. However, in contrast to distance relay protection, in the proposed algorithm the set of information features can be arbitrarily expanded. Fig. 7 illustrates the regions of  $\alpha$  and  $\beta$ -states in three-dimensional space, including the following information features:

- the resistance;
- the reactance;
- the amount of active power flowing through the protection.



**Fig. 7.** Location of training sample values in three-dimensional feature space

The error matrix typical for protection using a given feature space is given in Table 4.

**Table 4.** Error matrix corresponding to the modified k-nearest neighbors algorithm (k=15) in three-dimensional space

		Predicted class	
		$\alpha$	$\beta$
True class	$\alpha$	98,4 %	1,6%
	$\beta$	0%	100%

Due to the use of an additional information feature, the probability of recognizing  $\alpha$ -states increased to 98.4% completely excluding false operations. Thus, the three-dimensional algorithm of relay protection [1, 2], implemented on the basis of the k-nearest neighbors method, provides almost absolute state recognizability

whereas traditional relay protection (current and distance) have a low recognition ability.

#### 4 Conclusion

The implementation of distributed generation sources into electrical grids could lead to a decrease in the recognition ability of traditional types of relay protection. It is advisable to develop new organization methods of protection with high rates of technical perfection.

Simulation modeling based on the Monte Carlo method makes it possible to analyze a variety of potential states of the electrical grid operation in order to form relay protection algorithms that have the property of adapting to a specific circuit-state situation. The application of machine learning methods in the development of relay protection devices is promising.

The overhead line protection algorithm based on the k-nearest neighbors method identifies faults with a probability exceeding the distance relay protection abilities. In this case the complete absence of false operations of protection in operating and self-starting load states is provided.

#### References

1. Y.Y. Lyamets, G.S. Nudelman, D.V. Zinoviev, "Multidimensional relay protection, Part 1. Theoretical prerequisites", *Electricity*, no. 10, pp. 17-25, 2009.
2. Y.Y. Lyamets, G.S. Nudelman, D.V. Zinoviev, "Multidimensional relay protection, Part 1. Theoretical prerequisites", *Electricity*, no. 11, pp. 9-15, 2009.
3. I.V. Nagay, "The functioning provision of remote backup transformers protection in conditions of series-transverse asymmetry", *Izvestiya vuzov, North Caucasian region. Technical science*, no. 5, pp. 19-24, 2011.
4. S.M. Ermakov, "Monte Carlo Method in Computational Mathematics: Introductory Course", – St. Petersburg, P. 192, 2009.
5. C.M. Bishop, "Pattern Recognition and Machine Learning", – Berlin: Springer, P. 738, 2006.
6. Michie D, Spiegelhalter D, Taylor C. "Machine Learning, Neural and Statistical Classification",— Ellis Horwood, P. 290, 1994.
7. A.L. Gorelik, V.A. Skripkin, "Recognition methods: Textbook. Manual for universities", – M.: Higher School, P.222, 1977.
8. D.I. Bezdushniy, A.L. Kulikov, "The formation of generalized characteristics of the operation of relay protection based on the method of principal components", *Relay protection and automation*, Vol.34, no.1, pp. 20-27, 2019.
9. M.V. Sharygin, A.L. Kulikov, "Statistical methods for recognizing modes in relay protection and automation of power supply networks", *Power stations*, no. 2 (1039), pp. 32-39, 2018.
10. E.M. Schneerson, "Digital relay protection", – M.: Energoatomizdat, P.549, 2007.
11. G. Van-Tris, "The theory of detection, estimation and modulation. Volume 1: Theory of Detection, Estimation, and Linear Modulation: translation from english, ed. prof. V.I. Tikhonov", – M.: Soviet radio, P.744, 1972.