The support vector machine application in the implementation of multidimensional relay protection

A.L. Kulikov[1] , *D.I.* Bezdushniy[1,*], *M.V.* Sharygin[1] and *V.Yu.* Osokin[1]

[1]Nizhniy Novgorod State Technical University, Russia

Abstract. Power engineering digital transformation, the use of different intelligent electronic devices (IEDs), high-speed communication protocols provide extensive opportunities for relay protection and automation systems modernization of power utilities. One of the most promising avenues of power engineering development is design of new protection devices, whose principles are based on the elements of artificial intelligence and machine learning. The article discusses the features of the application of one of the most common machine learning algorithms, the support vector machine, by the example of constructing a three-dimensional fault detector, which would serve to increase a transmission line stepped protection selectivity. The proposed fault detector has high recognition ability and ease of technical implementation as part of the protection IED.

1 Introduction

Modern electric power systems are technically complicated complexes consisting of a large number of interconnected elements. At the same time, both producers and consumers of electricity impose increasingly stringent requirements for reliability of power supply. Relay protection plays an important role in the process of power engineering intellectualization. Modern relay terminals, in addition to the protection functions, quite often perform the functions of control, recording, oscillography, and a number of others. The exchange of information with other devices located in and outside the substation is provided by means of data acquisition and transmission systems [1]. In particular, the implementation of the IEC 61850 standard assumes the existence of a single intra substation process bus, not only for the relay protection devices, but also for other automation equipment [2]. This allows digital devices to access a large amount of information about the protected object in real time. Unfortunately, the basic algorithms of relay protection fundamentally unchanged over the past decades, and in fact they are digital analogues of their electromechanical predecessors [3, 4].

One of the most promising approach to the relay protection and automation organization is based on the multiple simulation and statistical processing of the simulation results [4 - 9]. In particular, it is advisable to use machine learning to implement the recognition functions of electric network states [10 - 12]. The use of this approach to power utilities protection potentially has the following advantages:

- features that help to make a decision on the protection operation can be individual for each object and be selected according to the condition of maximum information value. Since machine learning algorithms, as a rule, are multidimensional [3, 6, 7], it can be any number of features for a single protection limited only by the performance of a digital device that runs this algorithm;
- the choice of the parameters of the relay operating is connected with learning of the recognition algorithm via sample data characterizing the behavior of the protected object in normal and emergency states. Such data can be obtained both by the simulation modeling results and directly during operation. This approach allows protection to adapt to the operating conditions in real time (self-study).

Machine learning algorithms can be used not only as part of relay decision-making procedures, but also as a part of fault detectors that complement conventional types of relay protection and improve their main characteristics (speed, selectivity, sensitivity). Below is an example of the development of special fault detector for distance relay. It serves for the transmission line protection and it is based on one of the simplest and most illustrative but at the same time effective [11, 12] machine learning methods [6, 7] - support vector machine.

2 Principles of the support vector machine application

Problems solved by the machine learning can be divided into two main classes: classification and regression problems. The regression problem is to restore the form of some function $F(x)$ according to the available empirical information about the function values $F_1 ... F_n$ and the arguments of the function $x_1 ... x_n$. The

* Corresponding author: dimok.93@mail.ru

classification problem is to categorize some feature vector x as one of the classes $Y_1...Y_m$, based on a training sample consisting of a set of vectors $x_1...x_n$, which classes are known. The relay protection object is very close to the classification problem and can be formulated in machine learning terms. In this case, the subject of classification is the vector consisting of the parameters of currents and voltages calculated by measuring elements of protection device and the classes are the operating states of the protected object (normal operating state, short circuit in the protected zone, short circuit outside the protected zone, etc.).

The simplest form of the support vector machine is a two alternative classification method, the idea of which is to form a hyperplane in the feature space and separate elements related to different classes. Furthermore, a hyperplane should be as far as possible from representatives of both classes.

Suppose there is a training sample $\mathbf{x}_1 ... \mathbf{x}_N$, which consists of N vectors located in the M-dimensional feature space; then for each element of the training sample, there is a class label $y_1 ... y_N$ such that if the i-th element belongs to one class (we denote it by class α), then $y_i = 1$ and if the i-th element belongs to another class (we denote it by class β), then $y_i = -1$. Further, if the training sample is linearly separable, it follows that there is a hyperplane such that all the points on one side belong to the class α and all the points on the other side belong to the class β. Thus, we have the hyperplane equation

$$Q(\mathbf{a}) = \mathbf{w}^{\mathrm{T}} \cdot \mathbf{a} + b = 0, \quad (1)$$

where $\mathbf{w}$, b are coefficients that define the hyperplane; $\mathbf{a}$ is a vector describing an arbitrary point in the feature space where the hyperplane is constructed.

If the hyperplane really separates the training sample points, then we get

$$\forall i \in 1 \dots N,\ y_i \cdot \left(\mathbf{w}^{\mathrm{T}} \cdot \mathbf{x}_i + b\right) > 0. \quad (2)$$

In the general case, there is a hyperplane such that it satisfies expression (2) and it is not unique. The Fig. 1 shows the training sample points belonging to two classes and two options for the location of the separating hyperplane **a** and **b**. Since the considered example corresponds to two-dimensional space (Fig. 1), we see that the hyperplanes degenerate into straight lines. Despite the fact that both straight lines properly divide groups of points between themselves, it is clear from Fig. 1 that **a** is more suitable for the classification of the analyzed data because of providing a wide bandwidth between representatives of different classes. The width of this band is called the margin. The SVM method makes it possible to find a hyperplane that provides the maximum margin between classes.

Proven [11], that the equation of the hyperplane that provides the maximum margin can be obtained by finding a conditional minimum

$$\begin{cases} \min_{\mathbf{w}_m, b_m} \frac{1}{2} \mathbf{w}_m^{\mathrm{T}} \cdot \mathbf{w}_m, \\ \forall i \in 1 \dots N,\ y_i \cdot \left(\mathbf{w}_m^{\mathrm{T}} \cdot \mathbf{x}_i + b_m\right) \geq 1, \end{cases} \quad (3)$$

where $\mathbf{w}_m$, b_m are coefficients that define the hyperplane such that ensures the maximum margin.

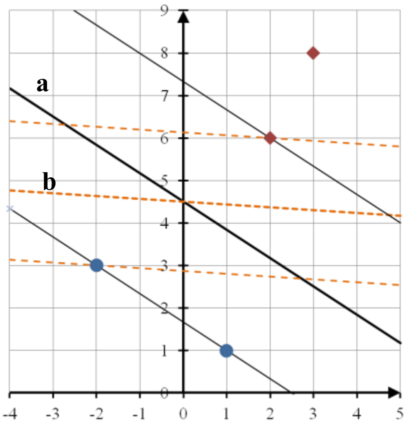

Fig. 1. Examples of separating hyperplanes.

However, in most practical cases, the training sample is not linearly separable. In other words, there is no hyperplane that satisfies a system of inequalities (2). For the SVM algorithm to be used in these cases, it is necessary to be modified in such a way as an incorrect classification of training sample objects is allowed. For the optimization objective, since there is a penalty for a total classification error, we have

$$\begin{cases} \min_{\mathbf{w}_m, b_m, \xi_1 .. \xi_N} \frac{1}{2} \mathbf{w}_m^{\mathrm{T}} \cdot \mathbf{w}_m + C \cdot \sum_{i=1}^{N} \xi_i \\ \forall i \in 1 \dots N,\ y_i \cdot \left(\mathbf{w}^{\mathrm{T}} \cdot \mathbf{x}_i + b\right) \geq 1 - \xi_i \ a, \\ \forall i \in 1 \dots N, \xi_i \geq 0 \end{cases} \quad (4)$$

where ξ_i is a slack variable; C is an error penalty.

In practice, the dual form of above optimization problem is widely used. Besides the fact that it gives the similar answer to the initial problem, it can be solved by faster iterative methods. In these terms, the optimization is performed with respect to the variables $\lambda_1 \dots \lambda_n$, corresponding to Lagrange multipliers. We say that the dual SVM problem for a linearly inseparable sample is based on the primal problem [7] and write

$$\begin{cases} \min_{\lambda_1 \dots \lambda_n} \left(\sum_{i=1}^{n} \lambda_i - \frac{1}{2} \cdot \left(\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \mathrm{y}_i \mathrm{y}_j \mathbf{x}_i^{\mathrm{T}} \mathbf{x}_{\mathrm{j}} \right) \right) \\ \forall i \in 1 \dots N,\ 0 \leq \lambda_i \leq C, \\ \forall i \in 1 \dots N,\ \lambda_i y_i = 0. \end{cases} \quad (5)$$

It is necessary to note that quadratic programming problem (5) can be solved by means of one of the well-known methods [for example, 9].

Since the solution gives the optimal Lagrange multipliers λ_1 ... λ_n (5), this means that we can recover the equation of the separating hyperplane

$$\mathbf{w} = \sum_{i=1}^{n} \lambda_i y_i \mathbf{x}_i, \qquad b = \frac{1}{y_s} - \mathbf{w} \cdot \mathbf{x}_s, \tag{6}$$

where s is the index number of the feature vector such that $0 < \lambda_s < C$.

Thus, using the coefficients λ_1 ... λ_n, we get the following classification algorithm

$$a(\mathbf{x}) = \operatorname{sign}\left(\sum_{i=1}^{n} \lambda_i y_i \mathbf{x}_i^{\mathrm{T}} \mathbf{x} - b \right). \tag{7}$$

Let us remark that for the correct operation of the classifier described by the expression (7), it is sufficient to sum over those i such that $\lambda_i \neq 0$. In other words, addition combines only support vectors. Note, that the number of vectors is considerably smaller than the total size of the training sample.

3 The support vector machine application in relay protection problems

Let us consider an example of using the SVM algorithm to increase the recognition ability of protection devices. Suppose a distance relay device is installed on a transmission of the electric grid section (Fig. 2). Suppose the primary zone of the relay protects the entire line $\omega 1$ and the backup zone is meant for protection of the adjacent lines ω_2 and ω_3. Besides, distance relay have to operate faults on transmission line ω_2 and ω_3 by separate zones, because it is necessary to provide required sensitivity and speed of a distance relay. Note also that zones have different sets and time delays and are denoted by "Zone ω_2" and "Zone ω_3". Now we implement the recognition algorithm for the distance relay such that it distinguishes faults on the lines ω_2 and ω_3.

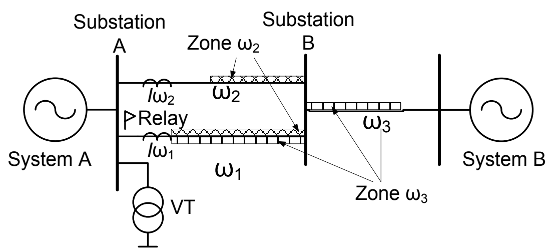

Fig. 1. Scheme of the analyzed power grid section.

Repeatedly simulating faults on the lines ω_2 and ω_3 in accordance with Monte Carlo method [10], we obtain the mode parameters (currents and voltages) at the protective device location in the form of complex values.

An analysis of the obtained model data shows that on the complex plane of active and reactive resistances, many short circuits on the lines ω_2 and ω_3 of different distances and with different values of the transition resistance will be located as shown in Figure 3.

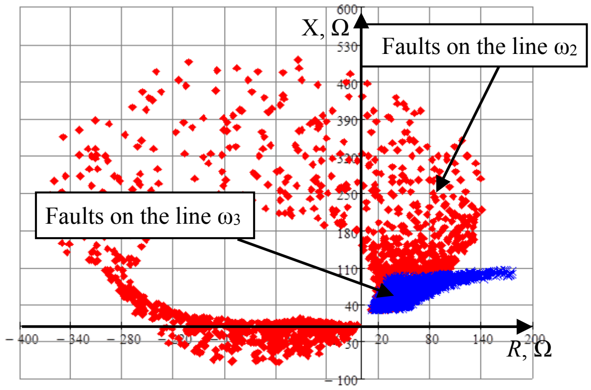

Fig. 2. The set of complex impedances in cases of different fault distances on the lines ω_2 and $\omega 3$.

According to Fig.3, practically all faults on line ω_3 are characterized by similar to line ω_2 fault resistances such that they are located in the "overlap" area, therefore, that they cannot be definitely identified. It is obvious that it is not possible to select the distance relay zone characteristics such that one of them ensures relay operation only at the line ω_2 faults and the other only at the line ω_3.

Let us use the support vector machines and design additional fault detector for operating at faults on the line ω_3 and blocking at faults on the line ω_2. Fig.4 shows the logic of the cooperative functioning of zone ω_2 and ω_3 fault detector and additional one.

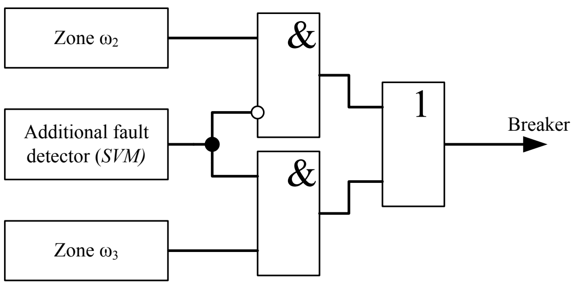

Fig. 3. The logic of the cooperative functioning of zone ω_2 and $\omega 3$ fault detector and additional one based on SVM.

Suppose that resistance (R), reactance (X), and the parallel line current ($I\omega_2$), as shown in Fig.2, are used as information features for implementation of an additional fault detector; then combine that features in a single three-dimensional space. The areas of existence of feature values in the resulting space are shown in Fig.5. Since the additional fault detector functions properly in conjunction with above zones (Fig.4), we see that it is possible to limit data points that are used as a training set by doing experiments such that at least one of these zones operates.

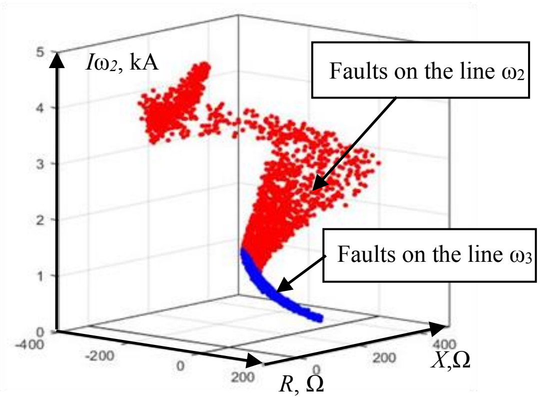

Fig. 4. The area of existence of fault points on the lines ω2 and ω3 in a three-dimensional space.

Such approach will increase the accuracy of the algorithm whereas learning time will be reduced.

Thus, the training sample consists of the data points we get in experiments such that faults occur in hatched areas of the line ω3 (shown in Fig.2), which we denote by α-states, and hatched areas of the line ω2 we denote by β –states. Fig.6 shows the training data points in the feature space.

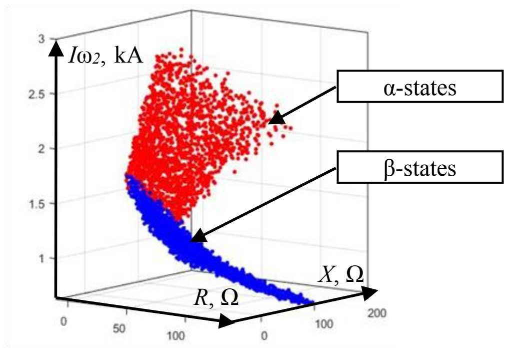

Fig. 5. The points in the training set

Let us find out how each of the selected features is effective for the classification problem solving individually. Fig.7 shows distributions of conditional probability density functions of the selected features in α- и β-states.

Based on each feature, we implement three one-dimensional classifiers. The classifiers will choose one of two hypotheses: "The fault on the line ω_2" and "The fault on the line ω_3" depending on the value of the corresponding feature. The decisive rule for the one-dimensional classifiers will be implemented according to the maximum likelihood estimation [for example, 14]. Each of the classifiers will choose the hypothesis which allows the obtaining feature value to be more likely.

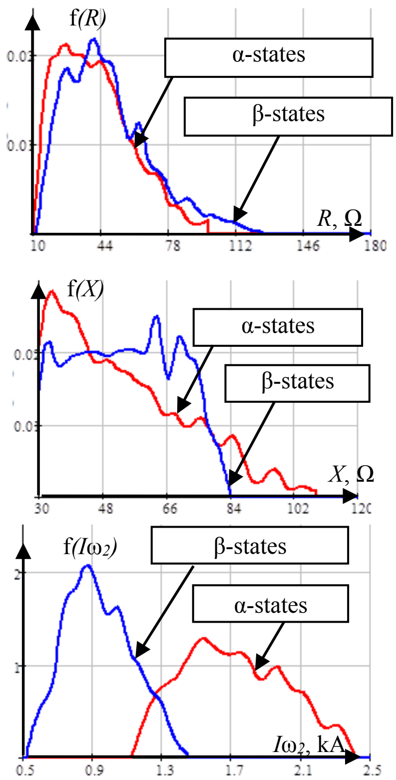

Fig. 6. Conditional probability density distribution functions of classification features.

Table 1 shows what error probability is provided by each of the features when the feature is used in the one-dimensional classifier. The probability distributions show that such features as resistance (R) and reactance (X) used alone provide probability of correct recognition almost the same as choosing the answer "at random".

Table 1. Error probabilities of one-dimensional classifiers.

R	***X***	***Iω2***
40,32%	41,21 %	4,79 %

Let us implement the "learning" procedure of the SVM algorithm with a linear kernel function [8] and the coefficient C = 100 on a previously obtained sample. The separating plane obtained as the result of the learning procedure is shown in the Fig.8.

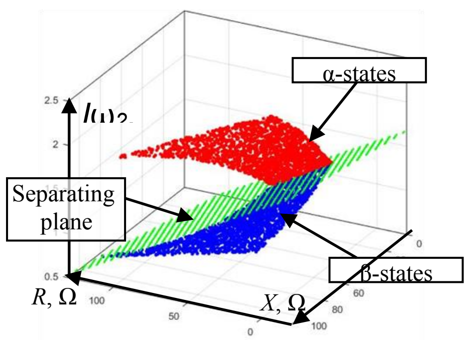

Fig. 7. Separating plane obtained as a result of the support vector machines.

Thus, the error percentage of the designed SVM-based recognition algorithm does not exceed 0.03% which is much less than the most effective one-dimensional classifier can provide. Therefore, the considered example confirms that the combination of information features into a single multidimensional feature space allows to achieve significant benefits in the task of relay protection.

4 Conclusion

The digitalization of power engineering and the development of modern information technologies made it possible to apply a fundamentally new approach to the relay protection designing, which based on simulation and machine learning.

The application of the support vector machine is promising in relay protection problems, both in the formation of new protection algorithms and in the use of SVM as an additional mean for increase the selectivity and speed of existing types of protection.

The combination of information features in a multidimensional feature space let us to implement a classifier such that it has a greater probability of the correct recognition than the use of information features.

This work was supported by the Ministry of Science and Higher Education of the Russian Federation (agreement No. 075-15-2019-1209 of May 31, 2019. Unique project identifier RFMEFI57717X0244).

References

1. I.N. Lizunov, A.N. Vasev, R.Sh. Misbakhov, News of Universities. Energetics problems. J. **1-2**, 52 (2017)
2. M. Adamiak, D. Baigent, R. Mackiewicz, Prot. Control J. **1**, 61 (2009)
3. E.M. Schneerson, *Digital relay protection* (Energoatomizdat, Moscow, 2007)
4. M.V. Sharygin, A.L. Kulikov, *Protection and automation of power supply systems with active industrial consumers* (NIU RANEPA, Nizhny Novgorod, 2017)
5. A.L. Kulikov, M.V. Sharygin, Electrical engineering J. **2**, 58 (2019)
6. D.I. Bezdushniy, A.L. Kulikov, Relay protection and automation J. **1,** 20 (2019)
7. M.V. Sharygin, A.L. Kulikov, Power stations J. **2**, 32 (2018)
8. H. Van Trees, *Detection, Estimation, and Modulation Theory: Detection, Estimation, and Linear Modulation Theory* (John Wiley & Sons, Inc., New York, 2001)
9. E. Gose, *Pattern recognition and image analysis* (Prentice Hall PTR, Upper Saddle River, 1997)
10. C.M. Bishop, *Pattern Recognition and Machine Learning* (Springer, Berlin, 2006)
11. L. Wang, *Support Vector Machines: Theory and Applications* (Springer, Berlin, 2005)
12. L. Bottou, C-J. Lin, Large Scale Kernel Machines J. **1,** 1 (2007)
13. S.M. Ermakov, *Monte Carlo Method in Computational Mathematics: Introductory Course* (St. Petersburg, 2009)
14. A.G. Zyuko, D.D. Klovsky, M.V. Nazarov *Signal theory* (Radio and communication, Moscow, 1986).