DEEP LEARNING IN RELAY PROTECTION OF DIGITAL POWER INDUSTRY

D.A. Stepanova, V.A. Naumov, V.I. Antonov
Chuvash State University, EKRA Research and Production Enterprise LTD
Cheboksary, Russia

ABSTRACT

Although modern relay protection exhibits all properties of an intelligent system, it has not yet fully acquired abilities to learn, adapt and recognize the modes of the protected electrical network. To give it these advantages, it is necessary to solve the central problem which is to distinguish the areas controlled by the protection of the electrical network modes parameters.

In relay protection all modes of network are divided into watched and alternative. In the first mode, the protection should be tripped, and in the second - the tripping is strictly prohibited. The problem of ensuring selectivity of protection can be considered at the rate of process as the establishment of belonging of the arriving data of an electrical network mode to a certain class in space of controlled parameters, i.e. having determined them to a class of the watched or alternative modes.

Traditional methods of relay protection learning are based on the application of the characteristics of operation differentiating the watched and alternative modes, thereby revealing some similarities with elements of the theory of artificial intelligence. At the same time the problem of finding acceptable operation border is solved with grace by methods of machine learning. Despite external similarity of schemes of algorithms, their main difference consists in a way of a task of characteristic of operation of protection. In traditional relay protection characteristic of operation is stored in the permanent memory, and in the protection with artificial intelligence - in the rewritable memory and is a part of neurons.

The paper presents a solution to the problem of differentiation of the electrical network modes on the basis of deep learning, considering the problem of formation of the tripping areas, for example, resistance relays as a definition of belonging to the relay measurements to certain classes in the space of controlled

978-1-7281-6381-9/19/$31.00 ©2019 IEEE

parameters. Algorithms of machine learning have universality, efficiency and allow approaching scrupulously the choice of characteristics of operation, using all the arsenal of intelligent classifier. As a result, intellectual relay protection gains ability to differentiate difficult untied areas of precedents (the modes of an electrical network) containing enclaves of the alternative modes. Besides, intellectual relay protection has a possibility to correct characteristic of operation in the conditions of its operation by training of neural network at new precedents. However, for this purpose the operational personnel have to give signs to new data, turning them into precedents. Thus, intellectual relay protection has ability to adapt to changes of an electrical network.

The paper covers mathematical foundations of the precedents sets separation of different modes of the electrical network on the example of the intelligent resistance relay using a support vector machine method. The advantage of the method consists in using the uniform principles to classify the modes of an electrical network both in case of linear, and in case of nonlinear separation of precedents. At the same time the solution of the linear separability problem is considered as a solution of the quadratic programming problem in the traditions of the theorem of Karush-Kuhn-Tucker. If it is impossible to recognize the parameters of a mode by a linear classifier, a nonlinear classifier is used, applying the mapping of the initial case space using special kernels to a higher-dimensional straightening space where the set of the modes becomes again linearly separable.

Possibilities of application of a support vector machine method to solve the problem of classification and deep learning of relay protection are shown. On the example of an intelligent resistance relay, the mechanism of the tripping characteristic adaptation to changes in network parameters is illustrated.

Keywords. Classification of the modes of the electrical network, learning of relay protection, machine learning, supervised learning, support vector machine.

INTRODUCTION

In relay protection all network modes are split into two groups. The first group comprises modes in which protection has to actuate and the second group comprises modes when actuation is absolutely forbidden. In this work [1] the first group is named watched groups and the second group is named alternative groups. In classic theory selectivity of relay protection is provided by proper selection of actuation parameters and this can be considered as its learning process. The task of protection learning is to provide it with possibility to classify electric network modes that implies differentiation of watched and alternative modes. From this point of view use of up-to-date methods of deep learning for protected electric network mode classification tasks and relay protection learning is appeared to be well founded [2, 3].

For this task we used support vector method that is known as Support Vector Machine – SVM [4, 5]. Advantage of this method is its capacity to define non-linear task of differentiation of precedent independent complex fields (electric network mode controlled parameters) in terms of square programming with limits in tradition of Karush-Kuhn-Tucker theorem [4].

THE STATEMENT OF THE PROBLEM

In traditional relay protection characteristic of operation that differentiate watched and alternative modes are generated according to results of calculations or simulation modelling of normal and emergency modes of protected electric network. Watched and alternative modes are expressed at plane of relay protection controlled parameters as points characterized with coordinates $\left(R_j, X_j\right)$. In the future we will provide measurements of relay protection (set of objects $\mathbf{X}$), related to the first class with feature $y_j = 1$, and to the second – $y_j = -1$. Object $\mathbf{x}_j$ with the appropriate feature y_j, will be called *precedent*. In other words relay learning is performed on set of precedents $(\mathbf{X}, \mathbf{y})$ – learning sampling:

$$\mathbf{X} = \left\{\mathbf{x}_1, \ldots, \mathbf{x}_j, \ldots, \mathbf{x}_n\right\} = \left\{\left(R_1, X_1\right), \ldots, \left(R_j, X_j\right), \ldots, \left(R_n, X_n\right)\right\},$$

$$\mathbf{y} = \left\{y_1, \ldots, y_j, \ldots, y_n\right\},$$

where $\mathbf{y}$ is vector of features of objects of set $\mathbf{X}$, $y_j \in \{-1, 1\}$.

GENERAL PROPERTIES OF METHODS OF RELAY PROTECTION AND ARTIFICIAL INTELLIGENCE

Protection selectivity provision task may be regarded as a definition at a pace of the process of belonging of the incoming data of the electrical network mode to a certain class in the space of controlled parameters.

In the task of building the characteristic of operation of traditional relay protection, elements of the theory of artificial intelligence are seen, if we consider the characteristic of operation as a tool in distinguishing between watched and alternative modes. For example, in the case of a resistance relay, the characteristic of operation in form of polygon (Figure 1) is generated by means of inequalities constraints:

$$\begin{aligned} &\left\langle \mathbf{w}_i, \mathbf{x}_j \right\rangle + w_{0i} \geq 0,\ \left(i = \overline{1,3}\right), \\ &\left\langle \mathbf{w}_i, \mathbf{x}_j \right\rangle + w_{0i} \leq 0,\ \left(i = 4\right). \end{aligned} \tag{1}$$

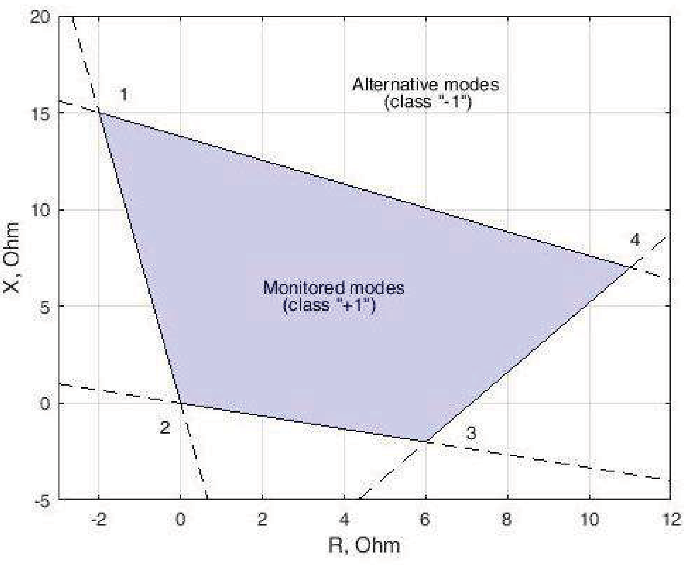


Figure 1 – Presentation of watched and alternative modes of electric network at the characteristic of operation of a resistance relay. Numbers of right lines correspond to equation numbers in inequality system (1)

The logic of the traditional relay is implemented in circuit that solves each inequality of system (1) in its channel and generates actuation signal as results of integration of all channel action (Figure 2, a).

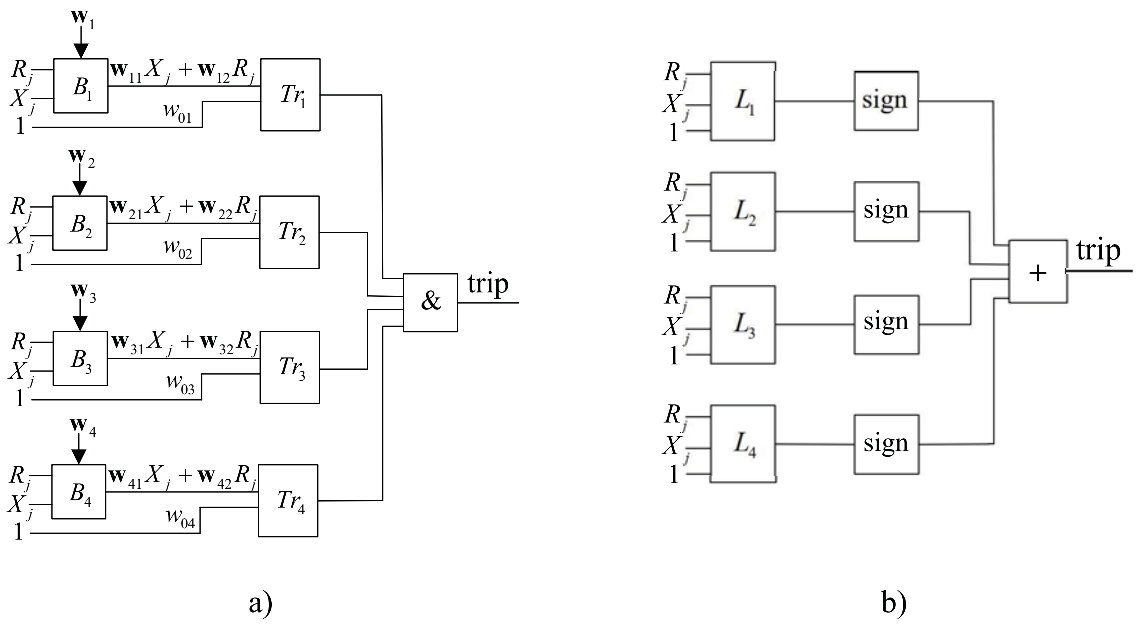
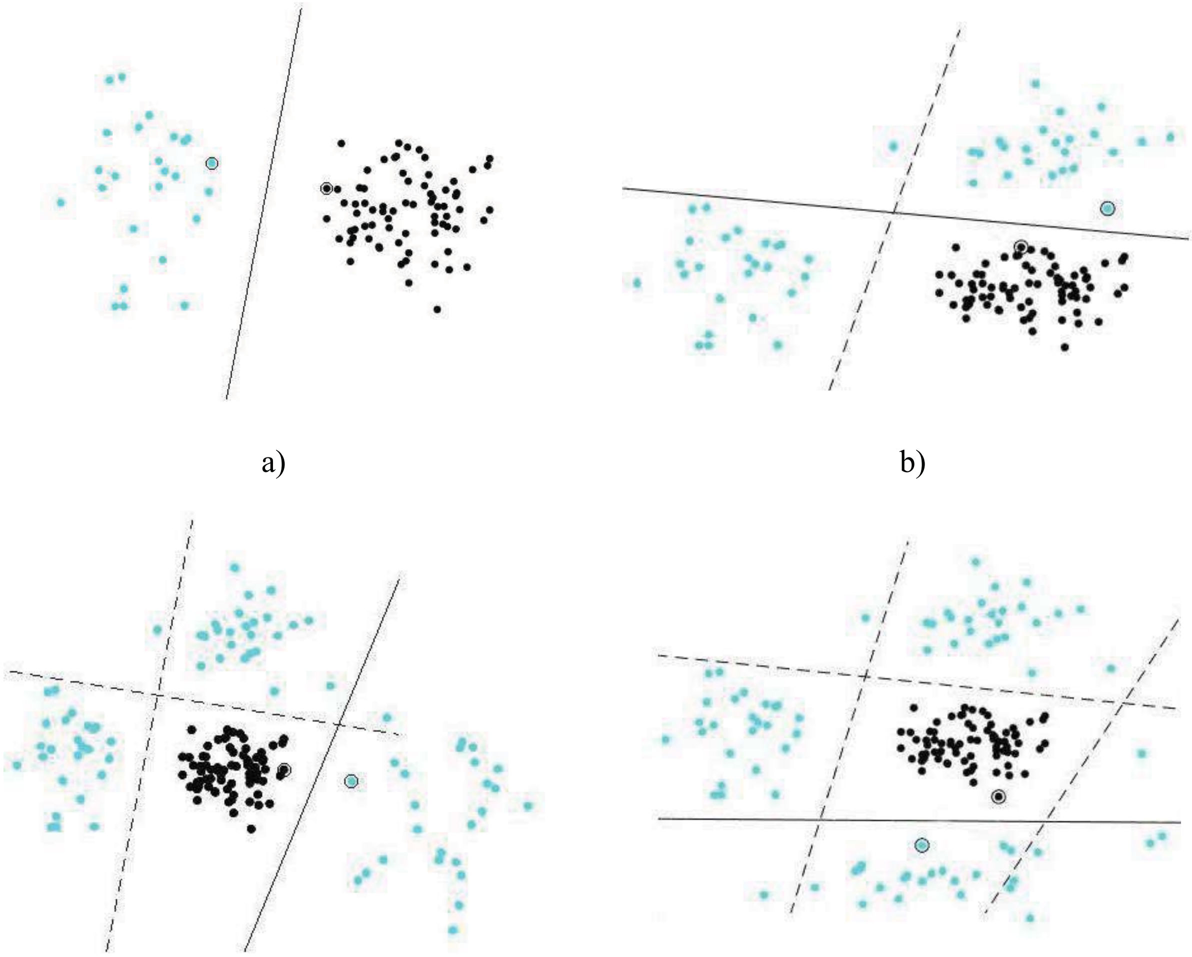


Figure 2 – Implementation of electric network mode differentiation algorithm in traditional relay protection (a) and artificial intellect based protection (b). Actuation parameters is generated with circuit (a) by operators of boundary line task Tr_i, circuit (b) – by neurons L_i, specifying differentiation line and blocks of function sign

The task of finding the required boundary of the tripping area of the relay can also be solved by applying machine learning methods. The algorithm for constructing the characteristic of operation (Figure 1) indicated at Figure 2 b demonstrates similarity with traditional relay protection algorithm. And, despite the external similarity of the algorithms, the main difference lies in the method of defining the characteristic of operation of the protection. In traditional relay protection it is set as weighting vector $\mathbf{w}_i$ in read-only memory and in artificial intelligence protection– in rewriteable memory and is part of neurons. This feature provides possibility of constant adaptation of intelligent protection to electric network changes.

The feature of learning to protect with artificial intelligence are demonstrated at Figure 3. As result of learning intelligent relay (Figure 2, b) will have the characteristic of operation indicated at Figure 4 a.

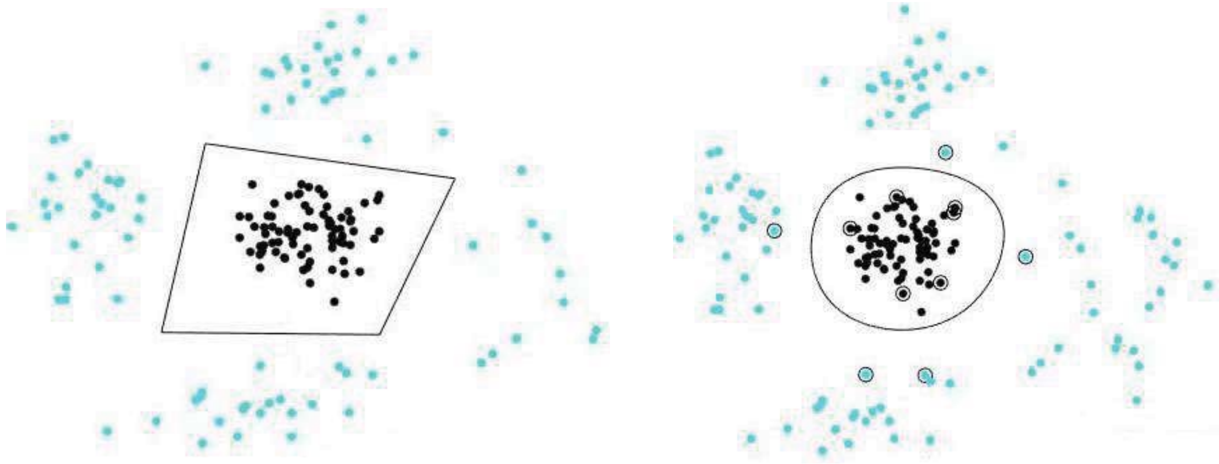


a) b)

c) d)

Figure 3 – Features of learning of the intellectual relay of resistance on the basis of linear classifier. The fundamental imperfection of the classifier is the inability to linearly divide the distributed precedent sets, which will require the participation of the teacher. Differentiation of precedents of sets of alternative mode distributed around precedents of watched modes is possible only as result of sequential learning of classifier [from a) to d)]. Dividing line of current stage is expressed as solid line and similar lines of previous stages – as dotted line. Dark points designate precedents of watched modes (class «+1»), light points – of alternative modes (class «-1») and circles – support vectors

Algorithms of machine learning are universal and permit to select meticulously the characteristic of operation using all possibilities of classifiers. As a result, intellectual relay protection acquires the ability to distinguish between complex unrelated areas. Thus, the characteristic of operation in the form of a polygon (Figure 4, a) for an intellectual relay based on linear classifiers will be formed with the inherent elegance using a nonlinear classifier (Figure 4, b).



a) b)

Figure 4 – Characteristics of intellectual resistance relays using linear (a) and non-linear (b) classifiers. Designations correspond to designations of Figure 3

Let's consider possibilities of using the method of support vectors for solving the problem of classification and deep learning of relay protection.

SUPPORT VECTOR METHOD IDEA

In terms of machine learning, the problem of forming triggering areas, for example, a resistance relay can be viewed as determining whether relays measurements belong to certain classes in the space of controlled parameters.

Consider the mathematical basis for the separation of sets of use cases of various modes of the electrical network on the example of an intelligent resistance relay using the support vector machine method (Support Vector Machine – SVM). It is required for the existing set of training precedents to determine the boundary of the relay triggering area, dividing class objects into use cases of monitored and alternative modes. The idea of the SVM method we consider the example of a linear separation of classes.

The purpose of the method is to build a classifier

$$a(\mathbf{x}) = \mathrm{sign}\left[f(\mathbf{x},\mathbf{w})\right], \quad (2)$$

returning a sign of a new object belonging to a certain class. Here $f(\mathbf{x},\mathbf{w})$ is discriminant function, $\mathbf{w}$ is classifier weighting vector. Discriminant function sign gives to object $\mathbf{x}_j$ feature of belonging to certain class y_j, converting it into precedent according to the following rules:

$$f(\mathbf{x}_j, \mathbf{w}) > 0, \text{ то } y_j = 1;$$
$$f(\mathbf{x}_j, \mathbf{w}) < 0, \text{ то } y_j = -1,$$

or in universal form:

$$y_j f(\mathbf{x}_j, \mathbf{w}) > 0. \tag{3}$$

Classifier obtains required properties as result of preliminary learning at precedent sampling $(\mathbf{X}, \mathbf{y})$. Task of learning is to find among all possible differentiating straight lines

$$f(\mathbf{x}, \mathbf{w}) = \langle \mathbf{w}, \mathbf{x} \rangle + w_0 = 0 \tag{4}$$

such one that is located at maximum possible distance from nearest precedents of both classes– *support vectors* $(\mathbf{x}_s, y_s)$. Here $\langle . \rangle$ is operator of scalar product, w_0 is scalar characterizing shift of differentiating straight line. Further this straight line will be called as *optimal differentiating straight line* [3] (Figure 5).

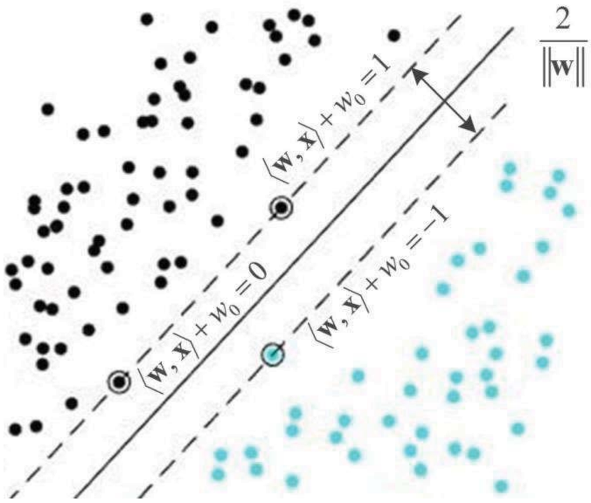


Figure 5 – Operation of linear classifier. Optimal differentiating straight line (solid line) is located exactly in the middle of boundary straight lines (dotted lines) defining differentiating band of maximum possible width. Designations are given on Figure 3

Boundary lines laid via support vectors of corresponding classes define differentiating band. Because differentiating band is not alone (Figure 6), then task of classifier learning may be defined as task of selection from set of differentiating bands of maximum width.

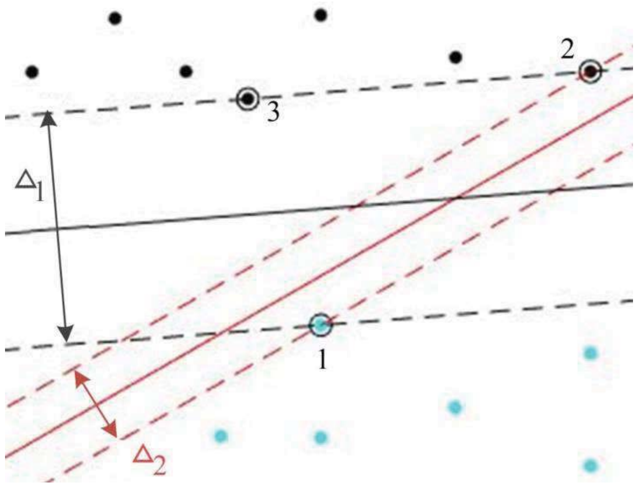


Figure 6 – Difference between optimal and sub-optimal differentiating bands. Two differentiating bands optimal (Δ_1) and sub-optimal (Δ_2) are shown. Boundary lines of both bands comprise support vectors 1 and 2, but optimal band considers also support vector 3 and has maximum possible width. All designations correspond to designations of Figure 3

Let us develop conditions of this task achievement.

Distance from differentiating straight line (4) to support vector $\mathbf{x}_s$ is defined with *indent*:

$$m_s = \frac{1}{\|\mathbf{w}\|}\left(\langle \mathbf{w}, \mathbf{x}_s \rangle + w_0\right) y_s = \frac{c}{\|\mathbf{w}\|}, \tag{5}$$

where $c = f(\mathbf{x}_s, \mathbf{w})$ is value of discriminant function in each point of corresponding boundary straight line.

Differentiating straight line does not change its position at linear scaling of coefficients $\mathbf{w}$ and w_0. Thus it is possible to ask for discriminant function at support vectors to assume values of features of the corresponding support vectors:

$$f(\mathbf{x}_s, \mathbf{w}) = y_s = \pm 1.$$

Then assuming that $c = 1$, we obtain new equitations of boundary straight lines:

$$f(\mathbf{x}, \mathbf{w}) = \langle \mathbf{w}, \mathbf{x} \rangle + w_0 = \pm 1. \tag{6}$$

Double indent defines band width:

$$\Delta = \frac{2c}{\|\mathbf{w}\|},$$

or taking into account that $c = 1$:

$$\Delta = \frac{2}{\|\mathbf{w}\|}.$$

This implies that minimum norm of weighting coefficient vector corresponds to maximum width of differentiating band:

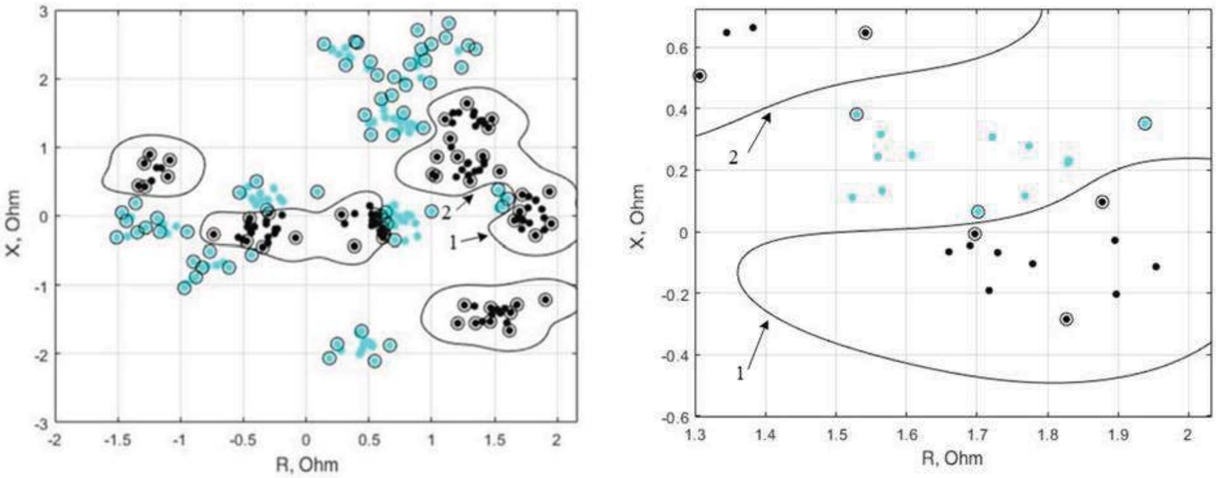
$$\|\mathbf{w}\|^2 \to \min_{\mathbf{w},\, w_0}. \tag{7}$$

Linear differentiation assumes that no learning sampling precedent is inside differentiating band and support vectors are located only at boundary straight lines.

INTELLECTUAL RELAY

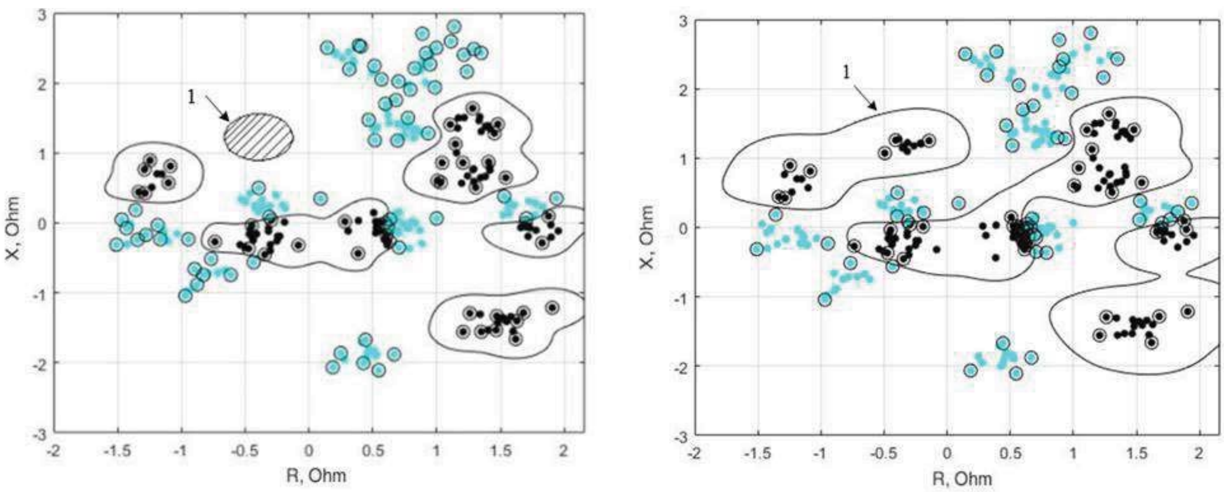
Intellectual relay is a neural network passing learning at different class precedent sampling. The configured network then classifies the new data that comes in real-time protection, relating them to one of the existing classes. The advantage of the relay learning method is its ability to find the optimal separating hypersurface in multidimensional space or the separating curve (line) on a plane in the case of both linear (Figure 5) and non-linear data separation (Figure 7).

At the same time intellectual relay has the ability to rebuild its characteristic of operation (Figure 8, a) during operation (Figure 8, b) learning the neural network at new precedents. For this purpose operating personnel shall give signs to new data, turning them into precedents.



a) b)

Figure 7 – The separation of areas of use cases watched and alternative modes of the electric network of an intelligence relay based on a nonlinear classifier. Figure 7 b demonstrates flexibility of SVM method at construction of differentiating lines. Designations are given in Figure 3



a) b)

Figure 8 – Separation of areas of use cases monitored and alternative modes of the electrical network using a nonlinear classifier in the mode of receipt of new data. When data appears in area 1, hatched in Figure 8a, the protection rebuilds the area of monitored modes again. Designations are given in Figure 3

LINEAR CLASSIFIER ALGORITHM

Let us consider optimum differentiating straight-line development algorithm.

As mentioned above to find optimum differentiating straight-line it is required to optimize criterion (7) at limits defining that precedents $\mathbf{x}_j$ of learning sampling $\mathbf{X}$ are outside of differentiating band field

$$y_j\left(\left\langle \mathbf{w}, \mathbf{x}_j \right\rangle + w_0\right) \geq 1. \tag{8}$$

Considered optimisation task is solved with Karush-Kuhn-Tucker theorem expressed as search task of Lagrange function saddle point:

$$\begin{cases} \mathrm{L}(\mathbf{w}, w_0, \boldsymbol{\lambda}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{j=1}^{n} \lambda_j \left[y_j\left(\left\langle \mathbf{w}, \mathbf{x}_j \right\rangle + w_0\right) - 1 \right] \rightarrow \min_{\mathbf{w}, w_0} \max_{\boldsymbol{\lambda}}, \\ \lambda_j \geq 0, \\ \lambda_j \left[y_j\left(\left\langle \mathbf{w}, \mathbf{x}_j \right\rangle + w_0\right) - 1 \right] = 0, \end{cases} \tag{9}$$

where L is Lagrange function, $\boldsymbol{\lambda}$ is Lagrange function.

Multiplier λ_j is defined positively. If precedent $\left(\mathbf{x}_j, y_j\right)$ is located at boundary straight-line

$$y_j\left(\left\langle \mathbf{w}, \mathbf{x}_j \right\rangle + w_0\right) = \pm 1,$$

then multiplier is strictly positive $\lambda_j > 0$ and for precedent equation holds true:

$$y_j\left(\left\langle \mathbf{w}, \mathbf{x}_j \right\rangle + w_0\right) - 1 = 0, \tag{10}$$

As mentioned above thus precedent is called *support vector*.

If precedent $\left(\mathbf{x}_j, y_j\right)$ is inner object of its class then multiplier $\lambda_j = 0$. Hence this precedent does not contribute to classifier learning and is called *periphery one*.

Solution of task (9) is narrowed to the same task of square programming in two stages. We first formulate conditions of obtaining of Lagrangian L extreme point in space of coordinates of vector of weight of classifier $\mathbf{w}$ and threshold w_0:

$$\begin{cases} \dfrac{\partial \mathrm{L}}{\partial \mathbf{w}} = \mathbf{w} - \sum\limits_{j=1}^{n} \lambda_j y_j \mathbf{x}_j = \mathbf{0}, \\ \dfrac{\partial \mathrm{L}}{\partial w_0} = \sum\limits_{j=1}^{n} \lambda_j y_j = 0. \end{cases} \tag{11}$$

At the second stage original optimisation task (9) is transformed in double task taking into account the following conditions (11):

$$\begin{cases} \mathrm{L}(\boldsymbol{\lambda}) = \sum\limits_{j=1}^{n} \lambda_j - \dfrac{1}{2} \sum\limits_{j,i=1}^{n} \lambda_j \lambda_i y_j y_i \left\langle \mathbf{x}_j, \mathbf{x}_i \right\rangle \to \max_{\lambda}, \\ \lambda_j \geq 0, \\ \sum\limits_{j=1}^{n} \lambda_j y_j = 0. \end{cases} \tag{12}$$

Type of discriminant function is characterised with parameters defined by system solution (12):

$$f(\mathbf{x}, \mathbf{w}) = \sum_{j=1}^{n} \lambda_j y_j \left\langle \mathbf{x}_j, \mathbf{x} \right\rangle + w_0,$$

where w_0 is defined with equitation (10):

$$w_0 = y_s - \sum_{j=1}^{n} \lambda_j y_j \left\langle \mathbf{x}_j, \mathbf{x}_s \right\rangle.$$

And $\lambda_j \neq 0$ is at support vectors only (at $j = s$, where s is support vector number) and $\lambda_j = 0$ at periphery precedents.

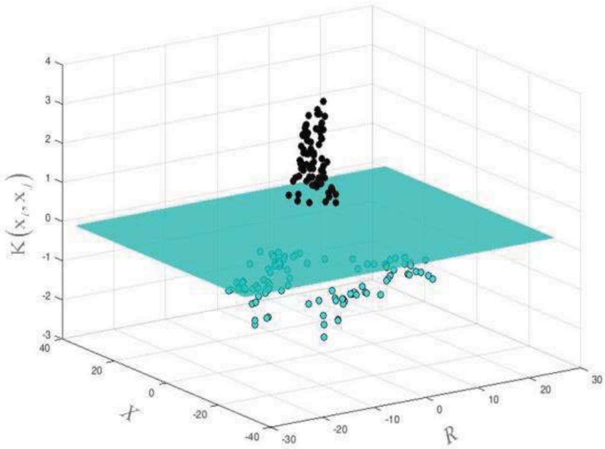
NON-LINEAR CLASSIFIER ALGORITHM

At classification of the most of real data linear selectivity of precedents is impossible. To recognize such data SVM-classifier which main feature comprises usage of special kernels is used. “Kernel trick” allows to express precedents from initial space to high dimensional space where set becomes linearly divisible [4]. Mathematical basis of this idea is Mercer theorem telling that if classes in some initial precedent space are not linearly divisible then this precedent space can be expressed in rectifying space where these classes will be linearly divisible.

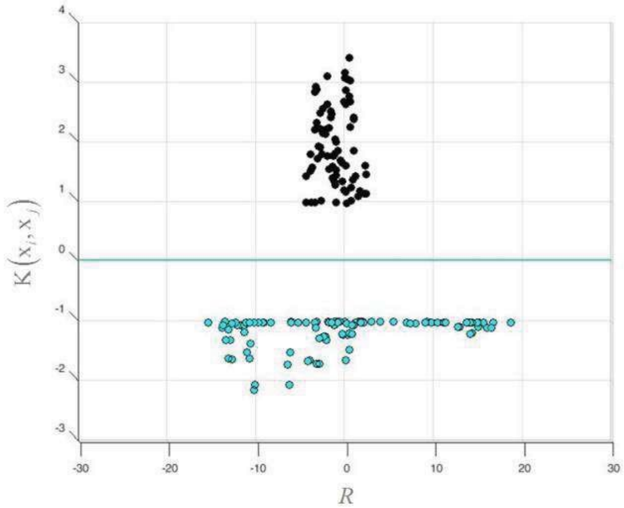
Precedents are expressed in new space using kernels–symmetrical functions

$$\mathbf{K}\left(\mathbf{x}_j,\mathbf{x}_i\right)=\left\langle \boldsymbol{\varphi}\left(\mathbf{x}_j\right),\boldsymbol{\varphi}\left(\mathbf{x}_i\right)\right\rangle, \tag{13}$$

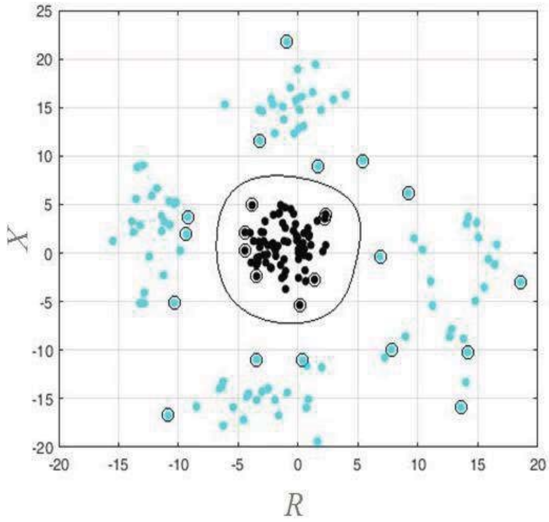
and SVM-classifier in new space will linear and its creation will be performed under the same rules as in case of classifier in precedent space $\mathbf{x}_j$ (Figure 9).



a)



b)



c)

Figure 9 – Rectifying space found by classifier. Precedents from original space (c) are expressed by means of function of kernel $\mathbf{K}(\mathbf{x}_j, \mathbf{x}_i)$ in new rectifying space (a), where they subsequently become linearly divisible (b). Figure 9 b demonstrates projection of rectifying space at plane (R, X). Designations are given in Figure 4

Figure 9 demonstrates transition of initial precedent space (Figure 4, b) to new rectifying space where precedents are linearly divisible. Transition to rectifying space is performed by replacing of scalar product $\langle \mathbf{x}_j, \mathbf{x}_i \rangle$ of Lagrangian function (12) by non-linear symmetric function $\mathbf{K}(\mathbf{x}_j, \mathbf{x}_i)$. It is important that no knowledge of function $\boldsymbol{\varphi}(\mathbf{x})$ of kernel is required because discriminant function is defined completely via kernel:

$$f(\mathbf{x}, \mathbf{w}) = \sum_{j=1}^{n} \lambda_j y_j \mathbf{K}(\mathbf{x}_j, \mathbf{x}) + w_0 .$$

Use in a great number of practical applications demonstrates high effectiveness of radial base kernel as Gaussian function:

$$\mathbf{K}(\mathbf{x}_j, \mathbf{x}_i) = e^{\frac{-\|\mathbf{x}_j - \mathbf{x}_i\|^2}{2\sigma^2}} = e^{-\gamma\|\mathbf{x}_j - \mathbf{x}_i\|^2} .$$

Example of non-linear differentiation of electric network mode precedents represented as independent fields is given in Figure 10.

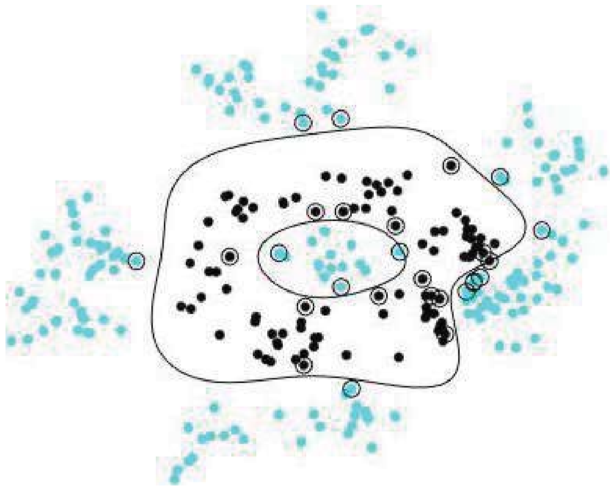


Figure 10 – Operation of non-linear classifier with radial function of kernel. Designations are given in Figure 3

SVM-classifier with radial base function is able to generate relay protection actuation parameters even if precedents of alternative mode are located in watched mode field.

CONCLUSIONS

1. Simulation modelling is main tool to define the characteristics of operation for both classic relay protection and machine learning methods.
2. Deep learning methods make possible to create relay protection complex the characteristics of operation with possibility to include alternative mode data enclaves in watched mode fields and reconfigure the characteristic of operation with learning of intelligence protection suing new precedents during operation.
3. Advantage of support vector method is in use of uniform principles for classification of electric network modes for both linear and non-linear divisibility.

LITERATURE

1. *Lyamets*, *Yu.Ya*. A Model of an Intact Power Facility As an Indicator of Damage / Yu.Ya. Lyamets, P.I. Voronov, M.V. Martynov, A.B. Atnishkin, M.Yu. Shirokin // Russian Electrical Engineering. – 2017. – Vol. 88. – No. 7. – 60-65 p.

2. *Stepanova, D.A*. Tasks of classification and deep learning in digital power engineering relay protection / D.A. Stepanova, V.A. Naumov, V.I. Antonov // RELAVEXPO-2019: Relay protection and automation of electric energy systems

of Russia: collection of proceedings of scientific-technical conference of young professionals. – Cheboksary. – 23-26 of April 2019. – 116-122 p.

3. *Stepanova, D.A*. To theory of relay protection deep learning / D.A. Stepanova, V.A. Naumov, V.I. Antonov // Dynamics of non-linear discreet electric and electronic systems: Materials of 13-th All-Russia scientific-technical conference // Chuvash State University Press. – Cheboksary. – 2019. – 319-327 p.

4. *Hastie, T*. The Elements of Statistical Learning, 2nd edition / T. Hastie, R. Tibshirani, J. Friedman // Springer. – 2009. – 533 p.

5. *Vapnik*, *V.N*. The Nature of Statistical Learning Theory / V.N. Vapnik // Springer. – 2000. – 314 p.