S.M. Dyachkov[1,2], V.A. Kuznetsov[1], V.N. Kutrunov[2], E.I.Yaroslavskaya[1]
1 - Tyumen cardiology research center, Tyumen, Russia
2 - Tyumen state university, Tyumen, Russia

UDK 004.891.3

EXPERIENCE OF APPLIANCE MACHINE LEARNING METHODS FOR MEDICAL DIAGNOSTICS

Abstract. The article specifies the formal definition of medical diagnostics problem. Negative and positive moments for solving problem of medical diagnostics by appliance different models such as logistic regression, artificial neural network, discriminant function and scoring model are described in the article. One of the possible schemes of appliance machine learning methods in clinic is performed in the manuscript.

Key words: medical diagnostics, logistic regression, artificial neural network, discriminant function, scoring model.

Formal definition of the medical diagnostics problem

The formal definition of the problem of constructing a diagnostic rule in medical science is pretty the same as the mathematical definition of the classification problem (the variety of machine learning tasks with the teacher) [1]. Let X is a finite set of features of objects (in medical science: potential signs of diseases or pathologies), Y - a finite set of numbers, names or labels of classes (in medical science: labels of diseases or pathologies). There is unknown target dependence $y^*: X \to Y$ where values Y are known only for the objects of the finite training sample $X^m = \{(x_1, y_1), \cdots, (x_m, y_m)\}$ (in medical science: $p_i = (\vec{x}, \vec{y})$ is a patient). It is required to find a function $F(X) = \overline{Y}$ capable of classifying a random object.

The formal definition and solution of the classification problem (the problem of constructing a medical diagnostic rule) depends on the feature space. A features space is called a mapping $f: X \to D_f$, where D_f is the set of admissible values of the features. The basic types of feature spaces that describe a vector of features are:

- Binary features: $D_f = \{0,1\}$;
- Nominal features: $D_f = \{0,1\}$ is finite set $x_j = \{x_{j_\alpha}\}$ where $\alpha = \overline{1,\eta}$ is given nominal value;
- Order features: D_f is infinite ordered set where $\forall x_{j_\alpha} \in x_j : \alpha_1 < \alpha_2 \rightarrow x_{j_{\alpha 1}} < x_{j_{\alpha 2}}$;
- Quantitative features: D_f is infinite set of real numbers $x \in [x_{\min}, x_{\max}]$, $x_{\min}, x_{\max} \in R$.

Based on the fundamental types of features more complex multidimensional feature spaces can be constructed, for example time series, images, video and many others.

In practice the following problems with dataset can occur: heterogeneous (different scales); incomplete (there are omissions); inaccurate (measured with inaccuracies); contradictory (the features of objects are the same, but the class label is different); excess (super-large dataset); insufficient (the number of objects a lot less than the number of features); unstructured (no structured data); non-trivial quality criteria.

That’s why all data have to be checked before starting problem-solving process. If some of dataset's problems can't be solved, thus, some limitations on the use of machine learning methods arise. When all problems with dataset are solved, we can start problem-solving process or developing medical diagnostic rule.

Problem-solving techniques

General statements.

Before starting developing the medical diagnostic rule it`s important to evaluate the classification function (medical diagnostic rule) by the loss function $L(a, x_i)$ used as the magnitude of the algorithm error $a \in A$ at the object $x_i \in X$. For the most problems loss function is used a quadratic error (1).

$$L(a, x_i) = (a(x_i) - y(x_i))^2 \tag{1}$$

Thus, empirical risk $Q(a, X^l)$ would be as the quality of the algorithm on the learning sample(2).

$$Q(a, X^l) = \frac{1}{l}\sum_{i=1}^{l} L(a, x_i) = \frac{1}{l}\sum_{i=1}^{l}(a(x_i) - y(x_i))^2 \qquad (2).$$

Fitting the model to training data is not always good for practical use. Very good fitting to training sample can be revealed because of noise, accidental emissions and various anomalies. All that improves the model's sensitivity, but the error of model for the unknown cases increases. It may happen because of an inadequate increase in the number of freedom degrees (model parameters). To avoid it there was modified the formula of calculation the empirical error. Modification can be divided in two groups: sample splitting and charging the penalty for big values.

Splitting sample method is called the "cross-validation" method. The main idea is dividing sample randomly on N subsamples and calculating mean error of a on subsamples (3)-(4).

$$X^L = X_n^l \cup X_n^k, L = l + k \qquad (3)$$

$$CV(\mu, X^L) = \frac{1}{N}\sum_{n=1}^{N} Q(\mu(X_n^l), X_n^k) \to \min \qquad (4)$$

Penalty charging method called regularization. The most common are Tikhonov (5) and Lasso (6) regularizations.

Tikhonov regularization

$$Q(a, X^l)' = Q(a, X^l) + \lambda \cdot \sum_j w_j^2 = \sum_i (a(x_i) - a(x_i))^2 + \lambda \cdot \sum_j \theta_j^2 \qquad (5)$$

where $\sum_j \theta_j^2$ - is the sum of the squared parameter's values of the classification function, and $\lambda \cdot$ is penalty size parameter selected manually;

Regularization of Lasso

$$Q(a, X^l)' = Q(a, X^l) + \lambda \cdot \sum_j |\theta_j| = \sum_i (a(x_i) - a(x_i))^2 + \lambda \cdot \sum_j |\theta_j| \qquad (6)$$

Another method is concerned the computing parameters of diagnostic (classification) rule. If the parameters of classification function will be computed by iterative algorithm, normalize input variables values is recommended. This transformation helps us to accelerate the process of convergence and, as a result, reduces empirical risk.

Let's examine some possible types of models suitable for diagnostic rules.

Discriminant function.

This is one of the simplest model, also called Fisher's linear discriminant function. The main idea is to find the best line that divide into the classes some objects by its features. The line is of the form (7).

$$f(\vec{x}) = p_0 + p_1 \cdot x_1 + ... + p_i \cdot x_i + ... + p_n \cdot x_n \quad (7)$$

After all parameters p_i is found by the optimization algorithm, the values of the discriminant function must be computed in the centroids of groups. After that for all class calculated non-intersecting, then intervals for the values of the discriminant function $a_i < f(\vec{x}) < a_{i+1}$ would be calculated corresponding to the particular class (8).

$$f(\vec{x}) = \begin{cases} y_1, f(\vec{x}) < a_1 \\ \quad ... \\ y_k, f(\vec{x}) \geq a_{k-1} \end{cases} \quad (8)$$

As we see the constructing a diagnostic rule using discriminant function not as easy as it seems and its quality dependence of many settings.

Logistic regression.

The most popular model for medical diagnostic rules [2]. It is associated with capabilities of the its parameters interpretation.

$$f(\vec{x}) = \frac{1}{1 + e^{-z}} \quad (9)$$

where $z = p_0 + p_1 \cdot x_1 + ... + p_i \cdot x_i + ... + p_n \cdot x_n$.

As shown by formula (9) the form of discrimination is still linear, but the area of intersection is stretched by logit transformation. It gives more accurate adjustments of cut-off point.

Logistic regression make it possible to estimate odds ratio of the i-th feature - $\mathrm{e}^{\mathrm{p_i}}$, it is very important for interpretation. For example, for feature with $\mathrm{e}^{\mathrm{p_i}} > 1$ increasing on 1 unit p_i involve increasing of probability $Y = 1$ on the $(e^{p_i} - 1)$ percent.

Scoring model.

Scoring models in practical medicine have become quite widespread due to the ease of their use and the lack of the need for the use of any computers and calculators. The most popular scoring models in medicine are the Apgar scoring scale [3], the GRACE scale for assessing the risk of developing the closest adverse cardiovascular outcomes (death, myocardial infarction) [4], as well as many other scales used in intensive care units [5]. Consequently, scoring models have proven themselves as a tool for rapid assessment of the condition of the examined patient.

General formula of scoring model is (10)

$$\mathrm{f}(\mathrm{x}) = \sum_{k=0}^{K} \mathrm{z}_\mathrm{k} \tag{10}$$

where z_k is a score for k-th feature calculated by (11).

$$\mathrm{z}_\mathrm{k} = \begin{cases} a_1^k, x_k < b_1^k \\ \quad \ldots \\ a_{n_k}^k, x_k < b_{n_k}^k \end{cases} \tag{11}$$

Scoring model is usually getting as rough logistic regression with weighted intervals. As a result, the error of scoring model will be bigger than error of logistic regression.

Artificial neural network.

A multilayer perceptron (MLP) is a kind of artificial neural network (ANN), which presented as a superposition of a large number of nonlinear functions obtained from a linear combination, that is reflected in formula (12) and Figure 1.

$$\begin{aligned} &MLP\left(\{W^k\}, X, \{\sigma^k\}\right) = \\ &= \sigma^K\left(\sigma^{K-1}\left(\cdots\left(\sigma^2\left(\sigma^1(X \times W^1) \times W^2\right)\cdots \times W^{K-1}\right)\right) \times W^K\right) = \vec{o} \end{aligned} \tag{12}$$

where $X : 1 \times n_0$, such as $n_0 = n + 1$, n - number of used features of the objects; $\{W^k\}$ - the set of matrixes of weights for each layer k, such as: $W^1 : n_0 \times n_1$ where n_1 - the number of neurons in the first hidden layer, $W^2 : n_1' \times n_2$, $n_1' = n_1 + 1$, n_2 - the number of neurons in the second hidden layer and so on to $W^K : n_{K-1}' \times n_K$, $n_{K-1}' = n_{K-1} + 1$ where n_K - the number of neurons in the output layer of MLP; $\{\sigma^K\}$ - a set of activation functions for each layer k [6]; $\vec{o}$ vector of MLP output signals.

As the parameters of MLP, one can distinguish [6]: the number of layers, the number of neurons in each layer, the activation functions in the layers, the error function.

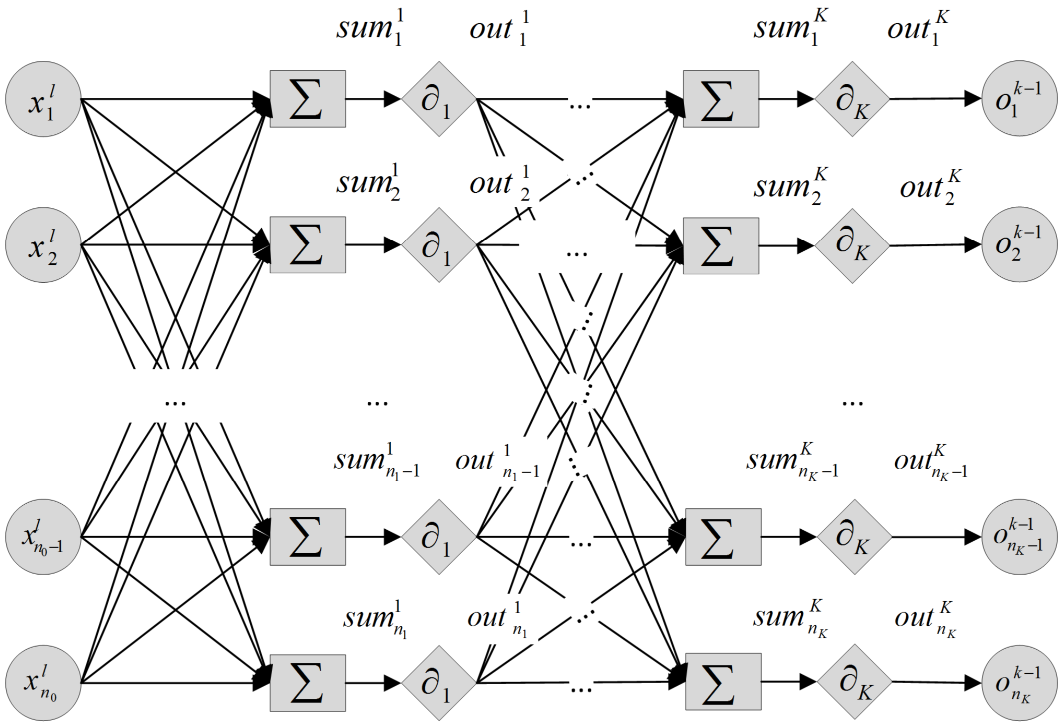

Fig. 1. Multilayer perceptron.

Source: designed by author

Practically MLP with two of three layers is sufficient for this purpose, since [1]:

- A two-layer ANN allows you to implement an arbitrary Boolean function.
- A two-layer ANN allows to define an arbitrary convex polyhedron.
- A three-layer ANN allows to separate an arbitrary polyhedral area, not necessarily convex, and not necessarily connected.

For evaluating the significance of the input vector's elements (features), the residual sum of square (RSS) can be used. RSS must be calculated for all the vectors of sample, but the for each input vector value of one of the element must be replaced by its average value in the training sample (13):

$$RSS_f = \sum_{l=1}^{L} \left(MLP\left(\{W^k\}, \{\sigma^k\}, \vec{x}_f^{\,l}\right)_i - y_i^l \right)^2 \qquad (13),$$

where $\vec{x}_f^l = \left(x_1^l, x_2^l, ..., \bar{x}_f, ..., x_{n_x-1}^l, x_{n_x}^l\right)$ is the modified input vector, in which the f-th attribute is replaced by the average value $\bar{x}_f$ in the training sample.

For value of $RSS_f < 1$ the f-th element can be excluded from the input value vector of the ANN, as it doesn't significantly affect the result of the classification.

Comparing of the problem solving techniques

According to the Table 1, we can say that the more complicated models is logistic regression and artificial neural network. However, in addition to the complexity of these models, we can also state their accuracy [2]. But for practical use they requires engineering calculator or a computer. The scoring model we obtained is easy to use. However, in comparison with logistic regression, discriminant function and artificial neural network, scoring model is less accurate in the overall diagnostic.

Table 1.

Models for medical diagnostics

Criteria	Discriminant function	Logistic regression	Scoring model	Artificial Neural Network
Analytic interpretation	Coefficients give information about way of relation (forward or backward) and its power		Coefficients can give information about relation power	No particular rule. Indirect assessment possible
Instruments needed to apply	Regular calculator	Scientific calculator	None	Computer
Type of classification board	Linear		Linear or non-linear (depend on model-building technique)	Non-linear

Source: based on [2]

Practice appliance of machine learning in medicine

Use of any of the discussed above methods in practice can be realized as a part of special system (external service) or as a module of the medical information system (internal service). Such systems and modules called as "clinical decision support system" (CDSS).

In general, CDSS consists of:

- Storage subsystem presented as a database of the diagnostic rules. Each diagnostic rule presented as PMML XML file [7].
- Execution subsystem presented as a web service of medical diagnostics.
- Graphical user interface for non-integrated usage presented as a web-site.

As shown on figure 1, connection through the internet to the CDSS from external medical information systems can be realized by using secured connection (as HTTPS) and depersonalization procedure (anonymization).

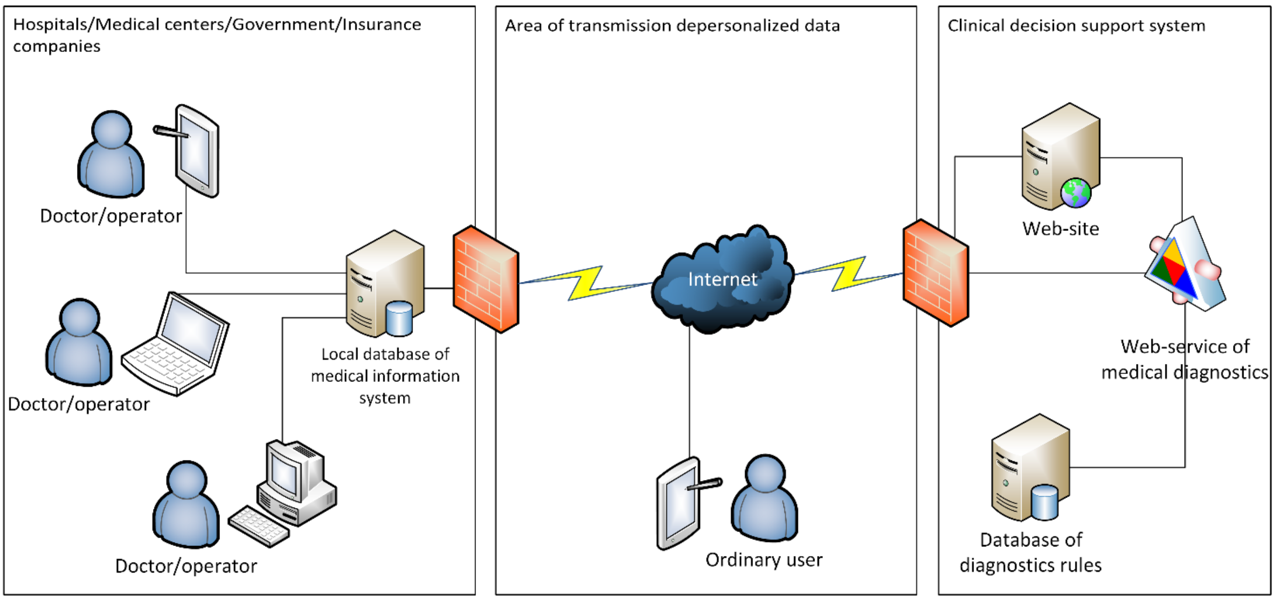

Fig. 2. Scheme of clinical decision support system based on machine learning diagnostics rules.

Source: designed by author

Conclusions

In this article all steps of applying machine learning methods for medical diagnostics from formal definition of the problem to practical use the method by clinical decision support system are described. In the research the problems of legalism, legal and ethical restraints of use such methods hadn`t been examined. This problem is of very complicated demand attention for lawyers and doctors. Another serious question is on which step such kind of medical diagnostics methods must be applied, and what quality of data (accuracy and validity time of input data etc.) must be.

Support

Financial support for this work was provided by the Federal Agency for Youth Affairs (Grant).

REFERENCES

[1] Vorontsov K.V. Mathematical methods of learning by precedents (theory of machine learning). - M. - 141 p.

[2] Kuznetsov V.A., Kutrunov V.N., Yaroslavskaya E.I., Dyachkov S.M. Statistical methods of medical data analysis in clinical practice // Scientific Review № 22 (2015) // P. 313-320.

[3] Sears W., Sears M. Your baby from birth to two years. - Moscow: Eksmo, 2012. - 912 p.

[4] Bassand J.-P., Hamm C. W., Ardissino D., et al. The Task Force for the Diagnosis and Treatment of Non-ST-Segment Elevation Acute Coronary Syndromes of the European Society of Cardiology. Guidelines for the diagnosis and treatment of non-ST-segment elevation acute coronary syndromes // Eur Heart J. №28(13) (2007) // P. 1598-1660.

[5] Vincent J-L, Moreno R: Scoring systems in the critically ill. Critical Care 2010, 14:207 [Electronic resource]. - Access mode: http://www.ccforum.com/content/14/2/207.

[6] Haykin S. Neural Networks: A Comprehensive Foundation / transl. from Eng. – M.: Publish house Williams, 2006. – 1004 p.

[7] Data mining group [Electronic resource]. - Access mode: dmg.org