INFORMATION TECHNOLOGY OF PROCESSING INFORMATION OF THE CUSTOMS CONTROL

Borys Moroz, Sergii Konovalenko

***Abstract**: The paper considers the issues and highlighted the development of information technology processing of information of customs control, which uses effective methods and tools to identify risks of violation of customs legislation. It was suggested take a holistic mechanism for the preparation, processing and interpretation of the results. The proposed information technology will increase the efficiency of the automated system of analysis and risk management Customs Service of Ukraine.*

***Keywords**: customs control, risk management, neural networks*

***ACM Classification Keywords**: I.5 Pattern Recognition – I.5.1 Models – Neural nets*

Introduction

Efficient operation of the Customs Service of Ukraine is possible only if the use of modern information technologies, methods and tools which will allow faster and better to solve the problem posed by the State [Пашко, 2008]. As is not uncommon to provide economic agents false information about the characteristics of goods for understatement (overstatement) of the customs value, transportation of smuggling, the use of high-quality business intelligence will improve the fight violation of customs legislation [WCO]. The introduction of such technologies is carried out in accordance with international standards and with the maximum involvement of the scientific and technical potential of the Customs Service of Ukraine, which allows you to select a part of the current research is the development and application of methods and tools for processing of information customs control.

Currently there are many methods and tools are used to develop decision support systems, classification, recognition, but from this mass want to highlight methods of artificial intelligence, model that use the neural networks. The theory of neural networks in the last decade gained a lot of practical applications in various fields of science and technology (face recognition, image, medicine and diagnostics, analysis stock and currency markets, using machines and robots). Such a wide distribution and availability of the many software packages modeling gives us reason to assume the possibility of using this mathematical model for the needs of the Customs Service of Ukraine.

Publications on the application of information processing methods of customs control and relatively few, mostly, they are conceptual in nature, highlighting the issues and relevance of the topic [Пашко, 2011]. In [Семенко, 2008] describes the possibility of applying the theory of fuzzy sets for the purpose of risk analysis, which allowed using membership functions to operate more flexibly with the calculation of risk. In our case, to ensure the quality of business intelligence is required to provide an opportunity to study and adapt the system that directs our choice in favor of neural networks.

Problem definition

The purpose of the article is to consider the theoretical and practical aspects of the development of information technology information processing customs control for risk analysis violations of customs legislation, in connection with which there is a need to solve the following problems:

1. To consider component parts of information technology;
2. To describe selected methods and means processing information of customs control;
3. To analyze effectiveness of the proposed system as a whole;
4. Based on the analysis to advice on possible application of the developed model.

Customs Risk Management – Economic security Ukraine

Development of world trade and foreign economic activity dictate their conditions to facilitate customs procedures during the import and export of goods, reducing of risks violation of economic security. In this regard, the implementation of customs emphasis shifted in favor of the application of risk management, the use of new technologies of nondestructive testing, close cooperation with customs services of other countries. This enables optimal use of available resources and the Customs Service to make the control of customs clearance in those areas where there is the greatest risk of violations, allowing the bulk of goods and individuals relatively free passage of customs control.

Consider in detail the system of risk analysis and management. Risk analysis violations of customs legislation is implemented using risk profiles (RP). RP applied during customs control and customs clearance of goods and vehicles crossing the customs border, and have a purpose to warn inspectors of a possible risk violations of the law during a specific foreign trade operations. Developing RP includes implementation of such actions:

- Definition of risk indicators;
- Selectivity rating;

Evaluation of the significance of this RP to fill budget;

Determine the effect of negative RP stories, etc.

The key concept in the customs risk management is a risk indicator. Indicator of risk - is a definite criterion that is used to identify potential violations of customs laws. This system, using the input data must carry out risk assessments on declaration or situation during customs clearance. If there is a risk, the system makes recommendations on the application of the necessary measures to minimize the damage from potential violation of customs legislation. Using a system of risk analysis methods and effective means of intelligent information processing will improve the quality of risk identification customs violations, smuggling.

System of intelligent information processing (SIIP)

Create an information technology processing information of customs control as noted above will use artificial intelligence. The process of developing the intellectual system of information processing (SIIP) generally requires the following steps (Figure 1):

- Perception of information from different sources (usually a heterogeneous vector of input data);
- Preparation or normalization of the input vector;
- Information processing;
- Interpetation of results.

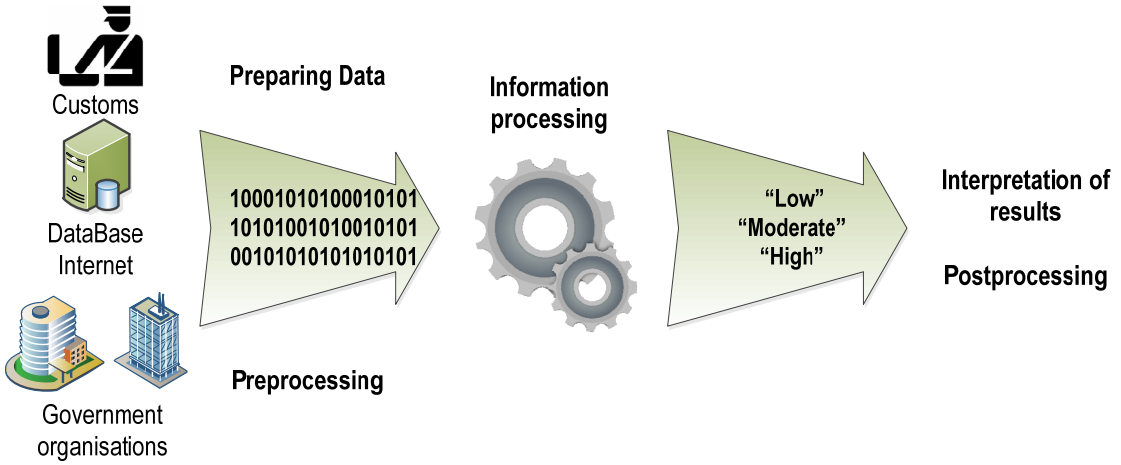


Figure 1. Model of the SIIP

Domain analysis of any problem means processing sufficiently large data sets, wearing a mainly polytypic nature and allocation factors and traits of the most important [Byuyul, 2005]. Since in most cases the information processing system receives input data for analysis from various sources, it becomes necessary to bring them to a suitable format for consideration. The primary source of data may make the storage and database of commercial and government organizations submitted documents, the internet network, i.e. much information as possible, which can be useful for decision making. Given the fact that intelligent systems have the properties of learning, it is important to pay attention to pretreatment and preparation of input data sets. Without doing this, we deteriorate the quality of information analysis system (pattern recognition, classification, etc.), and in some cases it will even fail to perceive the input vector data. The process of intelligent information processing depends on the choice of methods, which we believe meet the requirements to achieve the objective. Perhaps above all it is the methods using machine learning algorithms.

Finally, the interpretation of the results puts a goal before the reports in a clear formula for the user of the result of recognition or classification. In the context of risk analysis, violation of the customs legislation, this level of risk - "High", "Moderate" or "Low".

Preparing input vector "Information of customs control"

Preprocessing is the procedure for preparing data for analysis, during which they are in compliance with the requirements determined by the specifics of the problem being solved (subject domain) and used model processing (analysis) of the information received. Typically, data preprocessing includes two components [BaseGroup]:

1. Cleaning and optimization;
2. Transformation, normalization.

Cleaning is done in order to eliminate factors that reduce the quality of data and analytical algorithms hindering work. It includes processing duplicates, contradictions and fictitious values, restoration and filling gaps, smoothing and cleaning data from the noise suppression and editing of the abnormal value. In addition, the cleaning process restores the structure, completeness and integrity of the data, converted incorrect formats.

Optimization of data as part of pre-treatment includes reduction the dimension of the input data, the identification and elimination of non-significant features. The main difference between the optimization of purification, that the factors that are fixed in the cleaning process, significantly reduce the accuracy of the solution, or do the work of

analytic algorithms impossible. Problems to be solved by the optimization, data adapted to a specific problem and improve the efficiency of their analysis.

With regard to transformation and normalization of data, this step is required to make the information comprehensible to the terms used by the analytical model. This includes operations such as type conversion, quantization, encoding and so on. Each method of analysis requires that the original data were in any particular form. For example, neural networks work only with numeric data, and they should be normalized [Haykin, 1998]. Training data set SIIP must satisfy several criteria [BaseGroup]:

1. The representativeness - data should illustrate the true state of things in the subject area;
2. The consistency - conflicting data in the training set will result in a poor quality of the network training.

Table 1. Forming the input vector "Information of customs control"

№	Identification characteristics	Data types	Accepted values	Coding value
X_0	Country of Origin	string	Offshore	00 (bin)
			The EU countries	01 (bin)
			The EEA countries	10 (bin)
			other countries	11 (bin)
X_1	Product Code	integer	In accordance with the classifier	Range [0…1]
X_2	Customs cost	float	In accordance with the customs declaration	Range [0…1]
X_3	Quantity of goods	integer	Number of units or batches of delivery	Range [0…1]
X_4	Weight of goods	float	Weight unit of goods or the supply of the party	Range [0…1]
X_5	Invoice cost of a product	float	In accordance with the customs declaration	Range [0…1]
X_6	The difference gross and net product	float	no more 5%	0
			from 5% to 8%	0.5
			more than 8%	1
X_7	The history of the participant of foreign economic activity	string	black list	0
			gray list	0.5
			white list	1

These criteria provide for himself the whole complex of actions preprocessing of different types of input data. As a rule, neural networks like multilayer perceptron using sigmoid activation function. That is, the input vector $x = [x_0, x_1, \ldots, x_i]^T$ must be brought to the range [0...1] or [-1...1] by (1) or (2).

$$x_i = \frac{x_i}{\max(X)}. \tag{1}$$

$$x_i = \frac{x_i - \mu}{S}, \tag{2}$$

where S – range ($\max(X) - \min(X)$ or standard deviation, μ – average value.

If you need to partition the continuous value into segments of equal length, the quantization can be performed as the initial value of the division by a constant value (quantization step) and the integral part of the quotient:

$$y_q = \frac{y - y_0}{h}, \tag{3}$$

where h – quantization step.

An example of the process of forming the input vector is presented in the Table 1.

The components of the input vector are encoded by the following principle:

1. The variable X_0 is encoded binary numbers;
2. The variables $X_1 \ldots X_5$ are transformed by the formula (1) to the range [0 ... 1];
3. Variables $X_6 \ldots X_7$ take only three values, so we put them in compliance with three numeric values of {0, 0.5, 1}. These values match the level of risk {"Low," "Moderate," "High"} [ASYCUDA], [Konovalenko, 2012].

In such a way input vector was transformed to a common format and range – [0 ... 1], which is suitable for the activation function. Now it can be input into the neural network used for training and classification.

Information processing in customs control

System of intelligent information processing of customs control uses the methods and means of artificial intelligence. It was suggested [Moroz, 2011] the mathematical apparatus of artificial neural networks to detect the risk of violating the customs legislation. Quality training SIIP is probably the most important task, it is necessary to solve the developer, because it directly influences on the result [Bishop, 2007]. So first of all, you need the original set of input data separated on:

1. The training set $\left\{\left(x^{(1)}, y^{(1)}\right) \ldots \left(x^{(m)}, y^{(m)}\right)\right\}$;
2. Cross validation - $\left\{\left(x_{cv}^{(1)}, y_{cv}^{(1)}\right) \ldots \left(x_{cv}^{(m_{cv})}, y_{cv}^{(m_{cv})}\right)\right\}$;
3. Test - $\left\{\left(x_{test}^{(1)}, y_{test}^{(1)}\right) \ldots \left(x_{test}^{(m_{test})}, y_{test}^{(m_{test})}\right)\right\}$.

The distribution needs to be done in such proportions - respectively 60%, 20% to 20% of that total amount. This will allow us to use the $J_{train}(\Theta)$ (*Training error*), $J_{cv}(\Theta)$ (*Cross Validation error*) and $J_{test}(\Theta)$ (*Test error*). The general form of the error function (target function) has the following form:

$$J(\Theta) = \frac{1}{2m}\left[\sum_{i=1}^{m}\left(h_\Theta\left(x^{(i)}\right) - y^{(i)}\right)^2 + \lambda\sum_{j=2}^{n}\Theta_j^2\right], \tag{4}$$

where $h_\Theta(x) = g(\Theta^T x)$ (model that can be configured, or hypothesis), Θ - vector of parameters, which is configured, λ - parameter of regularization member.

Perceptron learning process gradient methods provides an iterative algorithm to minimize the target function $J(\Theta)$ by adjusting the parameters Θ (5).

$$\Theta_j = \Theta_j - \alpha\left(\frac{1}{m}\sum_{i=1}^{m}\left(h_\Theta\left(x^{(i)}\right) - y^{(i)}\right)x_j^{(i)} + \frac{\lambda}{m}\Theta_j\right), \quad (5)$$

where α - learning rate.

Learning rate α () must be chosen so as to avoid these problems [Ng, 2012]:

- If α too small, the gradient methods are quite slow to converge;
- If α too large, then it is possible that optimization process reusable "jump over" minimum of the target function, or even completely diverge.

During optimization, it is advisable to plot the target function $J(\Theta)$ of the number of iterations and stop the learning process when the decrease in the values of a single iteration will not be less than a certain threshold, for example 10^{-3}. Ideally, it depicts the relaxation process shown in Figure 2. b) During training SIIP appropriate and probably necessary to calculate the above errors ($J_{train}(\Theta)$, $J_{cv}(\Theta)$ and $J_{test}(\Theta)$) in the space of the corresponding sets. Analyzing these errors can be identified several issues that affect the ability of the classifier to generalize. These problems can be described by the following terms:

1) *Under fit* - λ has big value (Figure 2. a));
2) *Over fit* - $\lambda = 0$ (Figure 2. a));
3) *Just right* - λ has average value (Figure 2. a)).

The dependencies of the error function of the regularization parameter λ and capacity set *m* (Figure 3) reveals the insidious problem of over fit, i.e. situation where the classifier works correctly only on the training set.

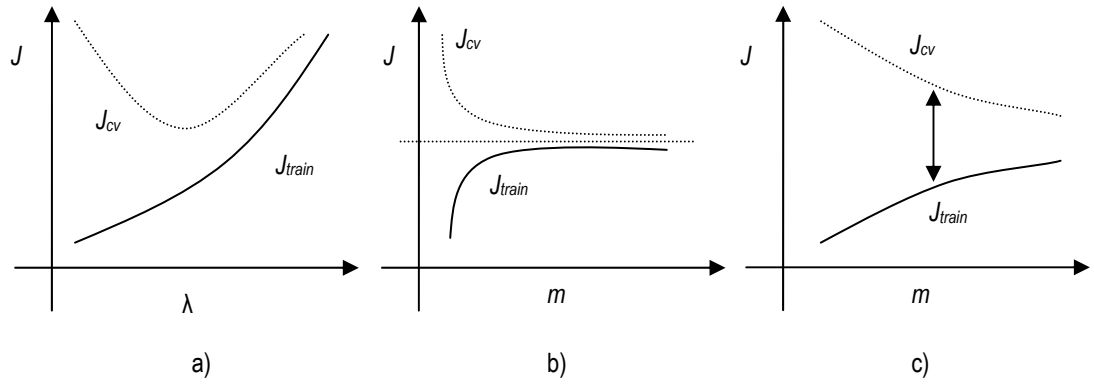


Figure 2. a) Dependence $J(\lambda)$, b) and c) Dependence $J(m)$, where *m* – capacity of set

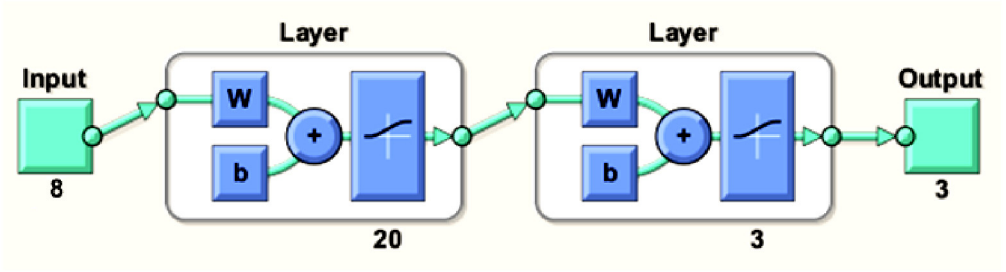
If during training and cross-validation we get a big mistake (Figure. 2. b), increasing the number *m* not solve the problem. In the situation (Figure 2. b)), the opposite - increase slightly improve the quality of education in general [Ng, 2012]. In addition, this procedure can add dimension reduction of the input space (if needed).

After spending 6 experiments for various architectures of multilayer perceptron, revealed that quasi-Newton method (BFGS) minimization of target function yields the Levenberg-Marquardt (LM) for neural network training

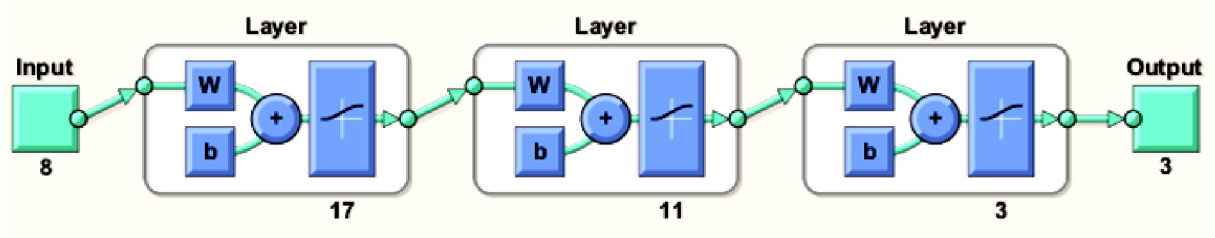
time and quality of recognition in the test set. Based on the results of experiments allocated optimal neural network model for one or two hidden layers:

1. Architecture with one hidden layer - 8-[20]-3, learning method LM, recognition error 1.4%;

Architecture with two hidden layers - 8-[17-11]-3, learning method LM, recognition error 1.2% (Figure 3 b)).



a)



b)

Figure 3. Neural model of the SIIP

So we got a complete description of effective models processing of information of customs control, which allow identifying the risk violation of customs legislation. And as a result will improve the quality customs clearance of goods, while reducing costs at the time.

Conclusion

As a result of this work was to develop a holistic model of risk identification violations of customs legislation on the basis of the type of neural network multilayer perceptron.

The paper discusses methods of converting continuous and discrete types of data suitable for analysis of the vector and sets. On the example domain "Information customs control" was shown how training set for the recognition of risk violations of customs legislation.

Were identified analysis techniques quality of the learning process SIIP and possible ways to overcome such problems of neural networks, as loss of the ability to generalize. Taking into account these recommendations improves learning algorithm and thus affect the quality of recognition risks violation of customs legislation.

On the basis of this experiments and comparative analysis of the quality of recognition of input images, highlighted the optimal architecture of the neural network.

Further research should be devoted to the study of effective methods of calculating the optimum multilayer perceptron architecture, as well as the mechanism of improvement opportunities neuroclassifier by sharing with fuzzy logic, using genetic algorithms.

Acknowledgement

The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA (www.ithea.org) and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine (www.aduis.com.ua).

Bibliography

[ASYCUDA] Automated SYstem for CUstoms DAta (ASYCUDA). Risk Management [Online]. Available: http://www.asycuda.org

[BaseGroup] BaseGroup Labs. Technology of data analysis [Online]. Available: http://www.basegroup.ru

[Bishop, 2007] Christopher M. Bishop. Pattern Recognition and Machine Learning (Information Science and Statistics). - Springer, 2007. - P. 738.

[Byuyul, 2005] A. Byuyul, P. Tsefel, SPSS: Art of information processing, (in Russian), DiaSoft, 2005.

[Haykin, 1998] S. Haykin, Neural Networks: A Comprehensive Foundation, 2nd ed., Prentice Hall, 1998.

[Konovalenko, 2012] Konovalenko Sergii. Preprocessing "raw" data sets as an important aspect of intelligent information processing, International Book Series "Information Science and Computing". Artificial Intelligence Driven Solutions to Business and Engineering Problems. - Rzeszow and Sofia : ITHEA, 2012. - Book 27. - P. 133-139.

[Moroz, 2011] B. Moroz, S. Konovalenko, "Aplication of neural networks within of the concept E-customs," (in Russian), in Information Science and Computing, book 22, Applicable Information Models, K. Markov, V. Velychko, Ed., 1st ed. Sofia, Bulgaria: ITHEA, 2011, pp. 104-110.

[Ng, 2012] Andrew Ng Machine Learning Online Course [Online]. Available: http://www.coursera.org.

[WCO] World Customs Organization. WCO course on risk assessment, profiling and targeting [Online]. Available: http://www.wcoomd.org

[Пашко, 2008] Пашко П.В. Основи митної справи в Україні: Підручник – К.: Знання, 2008. – 652 с.

[Пашко, 2011] Митні інформаційні технології: навч. посібник / О. Ф. Волик, О. В. Кащеєва, І. В. Дорда та ін.; За ред.

[Семенко, 2008] Семенко, О. М.-М. Підвищення ефективності роботи по протидії контрабанді та порушенням митних правил, митного контролю за рахунок впровадження автоматизованої системи аналізу та управління митними ризиками на основі Fuzzy - технології: до вивчення дисципліни / О. М.-М. Семенко. – К.: LAT & K, 2008. – 238 с.П. В. Пашка. - К. : Знання, 2011. - 391 с.

Authors' Information

Borys Moroz *– Professor, doctor of technical sciences, counselor of customs service of 1st rank, dean of faculty information and transport systems and technologies of the Ukrainian Academy of Customs, 2/4 Dzerzhinsky str., Dnipropetrovs'k, 49044, UKRAINE; e-mail: realtime07@mail.ru*

Major Fields of Scientific Research: information theory, fuzzy logic

Sergii Konovalenko *– inspector of customs service of 3rd rank, head of the laboratory of information systems and processes in customs affair of the Ukrainian Academy of Customs, 2/4 Dzerzhinsky str., Dnipropetrovs'k, 49044, UKRAINE; e-mail: customslab@rambler.ru*

Major Fields of Scientific Research: Computational neuroscience, computational bioinformatics