Neural net Decision Support System for Hand-written Author Identifications

A Ermolenko[1]

[1] Management Academy of the Ministry of the Interior of the Russian Federation, Moscow, Russia

Abstract. Questions of Decision Support System on processing of hand-written identifications are considered. The author has developed an original allocation method of unique handwriting characteristics on the basis of the line and symbols analysis. The informational content criterion is modified and parametric base divisibility is shown.

1. Introduction

In practice, police experts solve the problem of high-quality authentication of handwriting. Various complexes of automation of manual work of the expert criminalist are developed. However, the operation of these systems is focused on the operation of narrow professionals and is not designed for high-speed processing of large amounts of data. Handwriting authenticity assessment systems for large social systems, which not only automate the work of an expert, but also identify the author, are poorly developed. Advantages of fuzzy logic and neural networks are not fully used.

2. The control system model

Consider the property state protection system by police units. The main place in the tree of goals of police protection is the prevention of theft from protected objects and apartments (goal - A1). In addition, it is important to receive income from security services to the state budget (goal - A2), the disclosure of crimes (goal - A3), the search for criminals (goal - A4) and the prevention of administrative offenses (goal - A5). Formalize the model of management (1) of the tactical level protection unit in the form of [1]:

$$\langle S_o, T,\ Q, \perp S,\ A,\ B,\ Y,\ f,\ K, Y^* \rangle \ , \tag{1}$$

where: S_0 – problem situation; T – the time for making decisions; Q – available resources for decision-making; $S = (S_1, \ldots, S_n)$ – set of alternative situations that define the problem situation S_0; $A = (A_1, \ldots, A_K)$ – set of the goals required for decision making; $B = (B_1, \ldots, B_j)$ – set of limitations; $Y = (Y_1, \ldots, Y_m)$ – set of alternative decision; f – a function of the preferences of the decision makers; K – solution decision criteria Y^* – best decision [1].

The loss function (2) in the developed model is represented by:

$$f(\overline{x}) = (\max K(x), \min P(x), \min R(x))$$

$$\begin{cases} 0 < T < T_{\max} \\ 0 < Q < Q_{\max} \end{cases} \tag{2}$$

where K – income from security services (correlated with the number of protected objects), P –the risk of reimbursement from the Federal budget damage by theft or fire on protected objects, correlated with the introduction of coefficients α_i with the formalized values of the equipment of modern facilities by alarm systems *(P_1)*, by protection of building structures *(P_2)*, by authenticity and reliability of information on the functioning of the alarm system, security alarm control codes and authorized persons of protected objects *(P_3)* and other parameters of P_i forming:

$$P = \sum_{i=1}^{n} \alpha_i P_i \qquad (3),$$

R – Internal costs such as fuel costs *(R_1)*, ammunition for the training of rapid response teams *(R_2)* and other parameters collapsible additively taking into account the coefficients by β_j :

$$R = \sum_{j=1}^{m} \beta_j R_j \qquad (4).$$

Consider minimizing the risk of negative consequences of poor quality of security services, as well as reducing the internal costs of the organization [2]. Using risk management technologies with logic and probabilistic models (LPM) let's make L-scenario (figure 1) and identify 14 of the dangerous conditions:

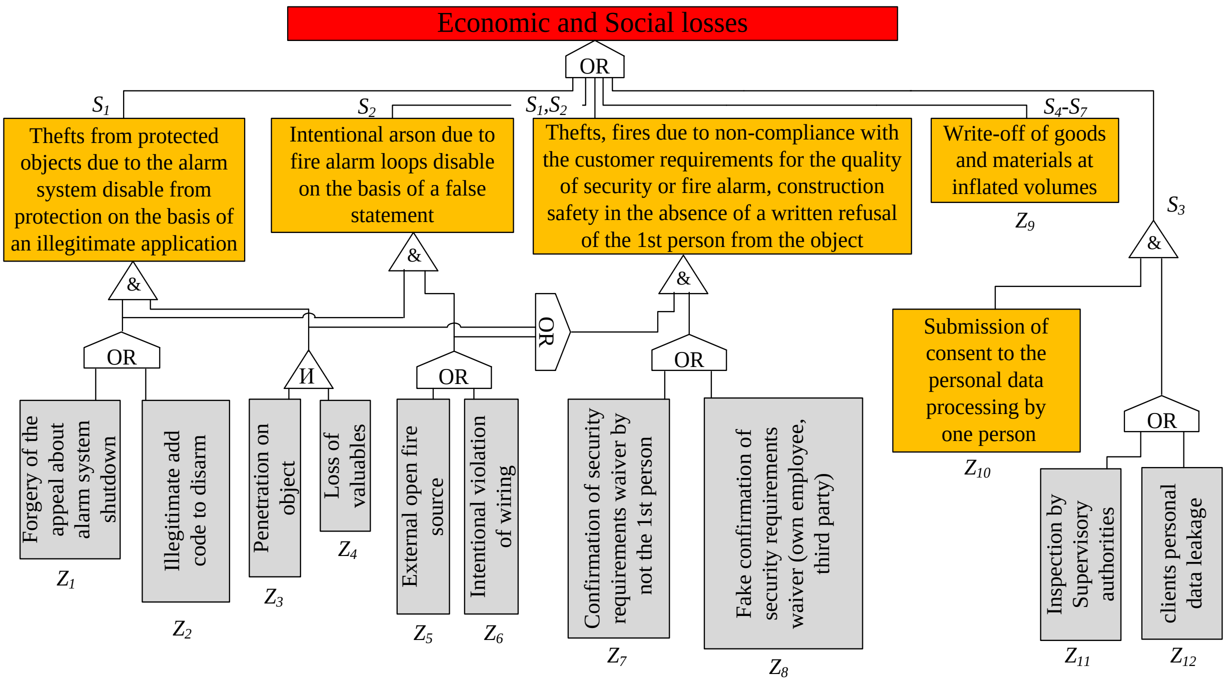


Figure 1. Logical scenario of dangerous States while ensuring safety of objects.

Given the above scenario, the l-function of the dangerous state is represented as:

$$y(z_1 \ldots z_{12}) = z_1 z_3 z_4 \wedge z_2 z_3 z_4 \wedge z_1 z_5 \wedge z_1 z_6 \wedge z_2 z_6 \wedge z_3 z_4 z_7 \wedge \\ \wedge z_3 z_4 z_8 \wedge z_5 z_7 \wedge z_5 z_8 \wedge z_6 z_7 \wedge z_6 z_8 \wedge z_9 \wedge z_{10} z_{11} \wedge z_{10} z_{12} \qquad (5)$$

We have studied the problem links of the formalized control model. The analysis showed that the actual aspect is the prevention of decision-making on forged handwritten documents. It is important to work with handwritten documents, not only customers but also their own employees.

The presented Cause and Effect Diagram as Fishbone Diagram by Kaoru Ishikawa [3] emphasizes that in the control process there is a need for expert support of decisions by means of artificial intelligence (figure 2). These tools should be focused on the development of new methods and models of handwriting identification.

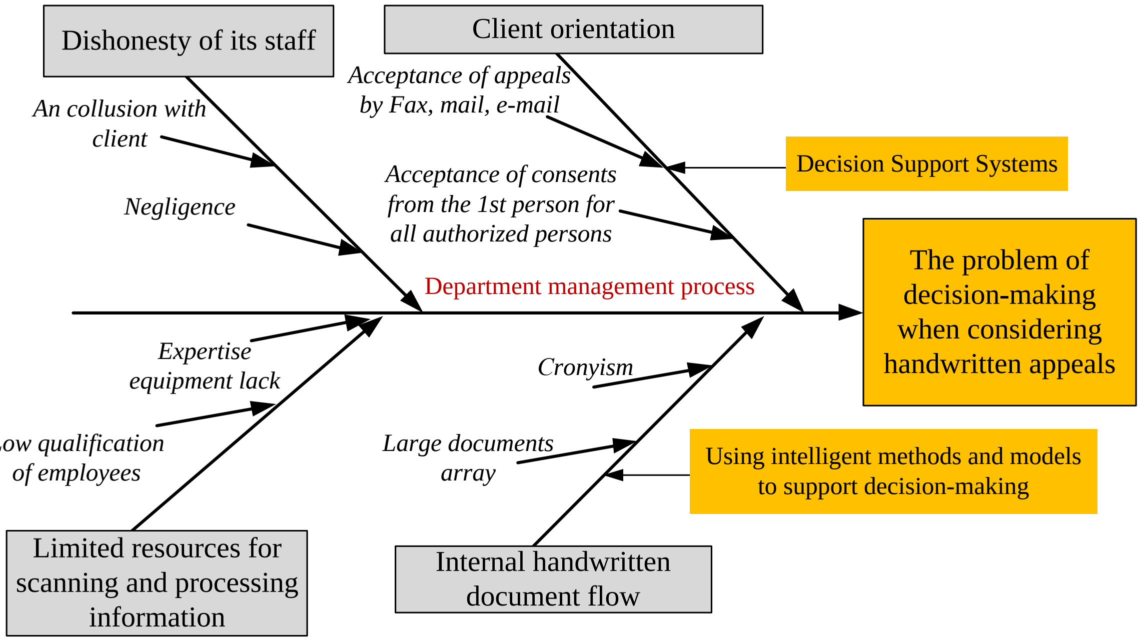


Figure 2. Causality of the decision-making problem.

3. Development of Decision Support System

The proposed decision support system (DSS) is based on fuzzy neural net identification (Figure 3).

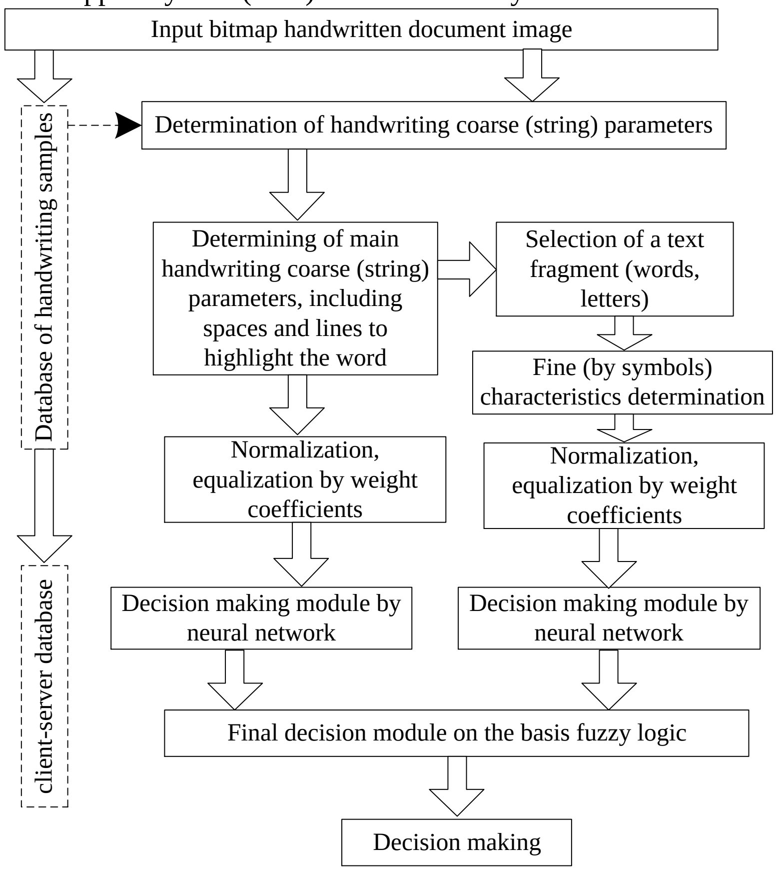


Figure 3. The proposed architecture of DSS.

The analysis carried out in the course of the research showed that different characteristics of the handwriting emphasize the uniqueness of its author in different ways [4]. In order to assess the informative characteristics of handwriting and their balancing, a modification of the Shannon criterion was developed (6):

$$\mu_j = \frac{\int\limits_{false\_min_j}^{t\_per_j} f\_false_j(x)\,dx + \int\limits_{t\_per_j}^{true\_max_j} f\_true_j(x)\,dx}{\int\limits_{false\_min_j}^{false\_max_j} f\_false_j(x)\,dx + \int\limits_{true\_min_j}^{true\_max_j} f\_true_j(x)\,dx} \tag{6}$$

Identification within groups of handwriting parameters is realized by means of neural networks. When creating the identification module, the possibilities of using various neural networks, including self-organizing, radial-basis and probabilistic ones, were studied and tested. At the same time, the studies have shown that the optimal decision model is the back propagation network. Formation of the training sample will produce formalization of handwriting characteristics of the original image. The solution classification problem of handwriting samples is solved in the tandem "original" - "falsification". Imagine the training and the test sample *F_O, T_O, F_T, T_T* differential form:

$$O = \begin{bmatrix} S_1 \\ S_2 \\ \dots \\ S_N \end{bmatrix} = \begin{pmatrix} x_1^1 - x_2^1 & x_1^2 - x_2^2 & & x_1^M - x_2^M \\ x_1^1 - x_3^1 & x_1^2 - x_3^2 & & x_1^M - x_3^M \\ \dots & \dots & & \dots \\ x_1^1 - x_N^1 & x_1^2 - x_N^2 & & x_1^M - x_N^M \\ x_2^1 - x_3^1 & x_2^2 - x_3^2 & & x_2^M - x_3^M \\ x_2^1 - x_4^1 & x_2^2 - x_4^2 & \dots & x_2^M - x_4^M \\ \dots & \dots & & \dots \\ x_2^1 - x_N^1 & x_2^2 - x_N^2 & & x_2^M - x_N^M \\ \dots & \dots & & \dots \\ \dots & \dots & & \dots \\ x_{N-1}^1 - x_N^1 & x_{N-1}^2 - x_N^2 & & x_{N-1}^M - x_N^M \end{pmatrix} \tag{7}$$

this will increase the sample size to

$$\dim(O) = M \times \frac{N(N-1)}{2}, \quad M = M_1 + M_2, \tag{8}$$

where M_1 – coarse (by strings) and M_2 – fine (by symbols) set of parameters.

To find the optimal algorithmic complexity balance and identification quality, various methods of backpropagation neural network training (gradient method, Lewenberg-Marquardt method, quasi-Newton method) has been investigated. The best convergence rate and minimize the mashing learning error $E_L = \sum_{X \in X} \|G(X) - Y\|$ and generalization error $E = \|G - F\|$ is been quasi-Newton BFGS algorithm (Broyden — Fletcher — Goldfarb — Shannon algorithm) [5]. Algorithm functioning is based on determination gradient vector at the *k*-th iteration g_k by multiplication $B_k P_k = -g_k$ that approximates Hessian-matrix by matrix B_k for next direction optimization P_k. *BFGS*-conversion formula is defined as

$$P_{k+1} = -g_{k+1} + \frac{(S_k, g_{k+1})Y_k + (Y_k, g_{k+1})S_k}{(Y_k, S_k)} - \frac{(S_k, g_{k+1})}{(Y_k, S_k)} \frac{1 + (Y_k, Y_k)}{(Y_k, S_k)} S_k \tag{9}$$

$S_k = X_{k+1} - X_k$ – vector equal to the change of X n the *k*-th iteration, $Y_k = g_{k+1} - g_k$ – the corresponding change of gradient.

In the BFGS–formula uses the results of the previous step. Thus, the 1st step should be carried out in a different way, for example, by the steepest descent method with minimizing the error of the neural network:

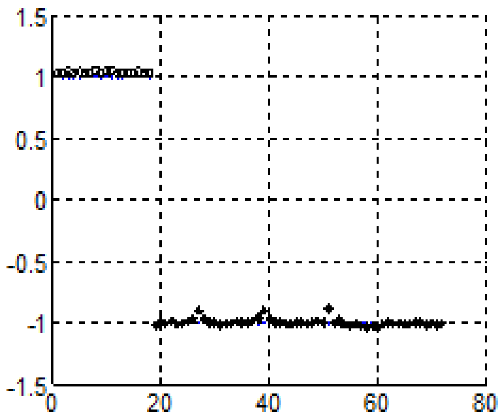
$$E(W)=\frac{1}{2}\sum_{j=1}^{p}\left(y_i-d_j\right)^2, \qquad (10)$$

where y_j – value of j-th output of neural network; d_j – desired value of j-th output; p – number of neurons in output layer. Changing the weight will produce the formula:

$$\Delta w_{ij}=-h\frac{\partial E}{\partial w_{ij}}, \qquad (11)$$

where h - the parameter that determines the learning speed.

The first layer of the proposed neural network topology represents 5 neurons with tangence-sigmoidal activation function. The second layer consists of 1 neuron with linear activation function. The number of training epochs was 2000. Research of the identifier model have shown (Figure 4) unambiguous separability of both string and character parameters at the output of the neural network.



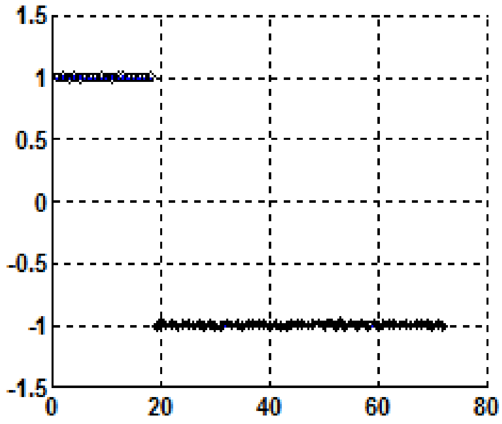


Figure 4. Neural network testing: left - for coarse identification, right - for fine identification.

Decision-making using fuzzy logic is realized on the basis of the composite rule by Mamdani inference by means of fuzzy implication:

$$\mu_C(z)=\mu_{C_1'}\vee\mu_{C_2'}=\left[\alpha_2\wedge\mu_{C_1}(z)\right]\vee\left[\alpha_1\wedge\mu_{C_2}(z)\right]. \qquad (12)$$

The input functions of the accessory are represented by bell-shaped distribution functions of characteristics $\mu_{X_i}(x)=e^{\frac{-(x-m_i)^2}{\sigma_i^2}}$. The linguistic output of the fuzzy model forms the answer about the authenticity of the handwritten document. It is possible to apply the centroid method (center of gravity method) of output characteristic defuzzification:

$$Z_{CT}=\sum_{j=1}^{n}\mu_C\left(Z_j\right)Z_j\Big/\sum_{j=1}^{n}\mu_C\left(Z_j\right) \qquad (13).$$

4. Identification of hand-writing characteristics

The coarse (string) characteristics determination of handwriting is realized by image segmentation, clustering [6], virtualization of incomplete strings clusters, two-level filtering and approximation of the reference trajectory writing of strings (figure 5).

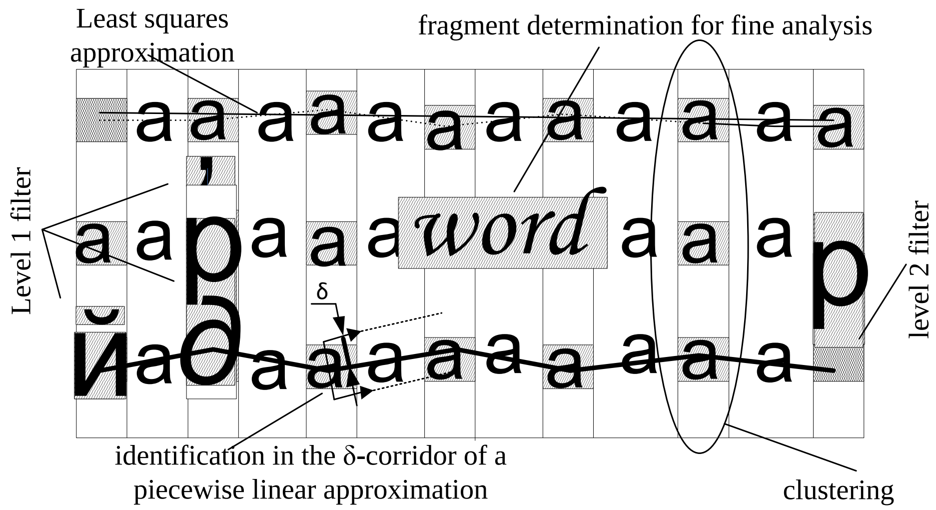


Figure 5. The text settings definition of coarse (row).

After filtration, it becomes possible to approximate the reference trajectory. Consider the clustering the image. We use the filter to remove small clusters of punctuation marks and large clusters of capital letters. Cut the pair strings joined in one cluster.

The approximation is feasible by the least squares method [7]. Differentiating the residual function by unknown parameters b and c and equating derivatives to zero, we obtain a system of equations:

$$\begin{cases} -\dfrac{1}{2}\dfrac{\partial x}{\partial b} = \sum\limits_{1}^{n}\left(y_i - bx_i - c\right)^2 x_i = 0 \\ -\dfrac{1}{2}\dfrac{\partial x}{\partial b} = \sum\limits_{1}^{n}\left(y_i - bx_i - c\right)^2 = 0 \end{cases} \tag{14}$$

The approximation of the reference trajectory allows forming an orthogonal virtual ruler:

$$\delta = h_{av} + \varepsilon \tag{15}$$

where h_{av} - the average clusters height, ε - the parameter of the symbols deviation, δ - the height of the virtual ruler.

The analysis of symbol pixel intensities when scanning with a virtual ruler allows to determine 10 coarse string parameters, as well as to select individual words for a thin (by symbol) analysis of the peculiarities of writing letters and inter-letter connections by the author (figure 6).

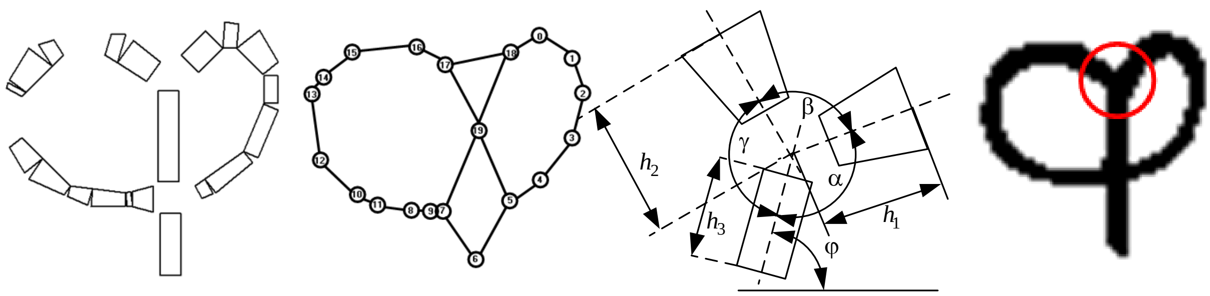


Figure 6. Character encoding by trapezoids and determination of 3-beam connections.

The developed fine symbolic identification methods is basing on coding of the symbols contour by trapezes [8] with dynamically changing bases and their subsequent filtration.

The developed method formation of vectors begins with the selection of points pair (for example, M and the next P) of the first studied curve of the handwriting string. Having the coordinates $\{x, y\}$ of the point, drawing a line through it $Ay + Bx + C = 0$ with deterministic parameters $\{A, B, C\}$. All points of contour $N(x_i; y_i)$ in the path between $\{M, P\} \forall Ax_i + By_i = C$ must meet the following condition:

$$|\delta| \le const, \quad \text{where } \delta = \frac{Ay_i + Bx_i + C}{\sqrt{A^2 + B^2}}. \tag{16}$$

Determination is realized by consecutive description of the symbol in the graph representation with formation of statistical matrix and adjacency list as well as auxiliary list (figure 7).

```
    0 1 2 3 4 5 6 7 8 9 10111213141516171819
 0 |.|X|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|X|.|
 1 |X|.|X|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|
 2 |.|X|.|X|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|
 3 |.|.|X|.|X|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|
 4 |.|.|.|X|.|X|.|.|.|.|.|.|.|.|.|.|.|.|.|.|
 5 |.|.|.|.|X|.|X|.|.|.|.|.|.|.|.|.|.|.|.|X|
 6 |.|.|.|.|.|X|X|X|.|.|.|.|.|.|.|.|.|.|.|.|
 7 |.|.|.|.|.|.|X|.|.|X|.|.|.|.|.|.|.|.|.|X|
 8 |.|.|.|.|.|.|.|.|.|X|.|X|.|.|.|.|.|.|.|.|
 9 |.|.|.|.|.|.|.|X|X|.|.|.|.|.|.|.|.|.|.|.|
10 |.|.|.|.|.|.|.|.|.|.|.|X|X|.|.|.|.|.|.|.|
11 |.|.|.|.|.|.|.|.|X|.|X|.|.|.|.|.|.|.|.|.|
12 |.|.|.|.|.|.|.|.|.|.|X|.|.|X|.|.|.|.|.|.|
13 |.|.|.|.|.|.|.|.|.|.|.|.|X|.|X|.|.|.|.|.|
14 |.|.|.|.|.|.|.|.|.|.|.|.|.|X|.|X|.|.|.|.|
15 |.|.|.|.|.|.|.|.|.|.|.|.|.|.|X|.|X|.|.|.|
16 |.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|X|.|X|.|.|
17 |.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|X|.|X|X|
18 |X|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|X|.|X|
19 |.|.|.|.|.|X|.|X|.|.|.|.|.|.|.|.|.|X|X|.|
```

```
  0:  1, 18,
  1:  0, 2,
  2:  1, 3,
  3:  2, 4,
  4:  3, 5,
√ 5:  4, 6, 19,
√ 6:  5, 6, 7,
√ 7:  6, 9, 19,
  8:  9, 11,
  9:  7, 8,
 10:  11, 12,
 11:  8, 10,
 12:  10, 13,
 13:  12, 14,
 14:  13, 15,
 15:  14, 16,
 16:  15, 17,
√17:  16, 18, 19,
√18:  0, 17, 19,
√19:  5, 7, 17, 18,
```

```
   5:  4, 6, 19,
   6:  5, 6, 7,
   7:  6, 9, 19,
1  17:  16, 18, 19,
2  18:  0, 17, 19,
   19:  5, 7, 17, 18,
```

Figure 7. Matrix, the adjacency list; auxiliary list.

On the basis of the proposed model of decision-making on neural networks and identification methods of 14 handwriting characteristics we have developed software. The main window of the developed software is shown in figure 8.

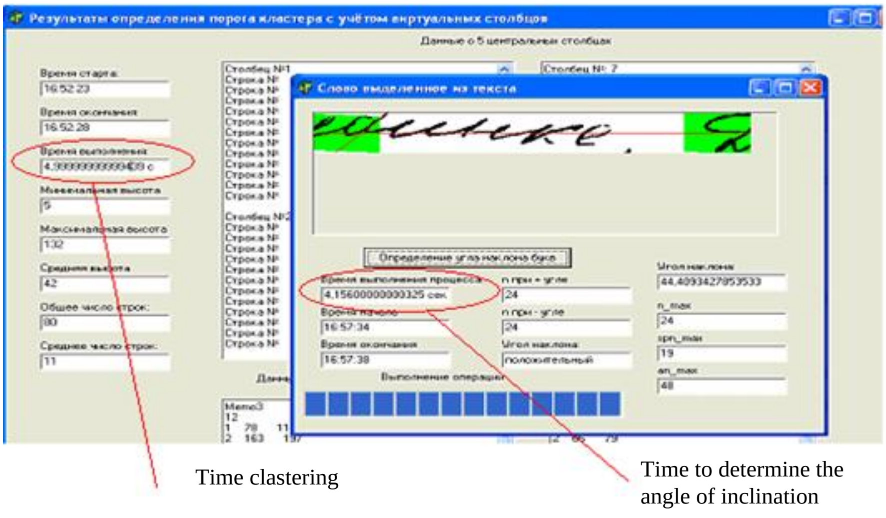


Figure 8. The main Window of the developed software DSS.

The proposed methods make it possible to formalize 14 handwriting characteristics, with their subsequent inclusion in the DSS. The evaluation of information content allows us to conclude that the most informative character characteristics are the angles of 3-beam connections, and from the number of string characteristics - the relative height of the letters and the angles of inclination. The time of operation (t) of the DSS consists of the work of string (t_1) and character (t_2) neural network identification, as well as generalizing fuzzy inference (τ)and is not more than 7 seconds:

$$t = t_1 + t_2 + \tau \leq 3\sec + 3,5\sec + 0,5\sec \leq 7\sec \quad (17)$$

The proposed algorithms were tested on an independent sample of 1000 samples. Confidence interval a positive identification in this case was 0.9.

The software has received state registration. The software is implemented in the client-server database concept and can be integrated into law enforcement databases.

5. Conclusion

As a result of researches the decision support system allowing identifying the author of the hand-written text is synthesized. The decision model based on BFGS-neural network with the method of training on the backpropagation algorithm is selected. The developed fuzzy-neural network model of identification and methods of information processing and determination of the unique characteristics of handwriting are implemented in ready-made software that allows identifying the author of handwriting within 7 seconds of computer examination with a confidence interval of 0.9.

6. References

[1] Solojentsev E 2004 Scenario Logic and Probabilistic Management of Risk in Business and Engineering. *Springer*.

[2] Solojentsev E 2010 Logical and Probabilistic Risk Management invalidity flight tests of machines, processes and systems *// Problems of Risk Analysis, Volume 7*, N4. pp 72 – 85

[3] Solojentsev E and Lebedev Y 2006 Logic-probabilictic models of company management unsuccess risk / *Mathematical Economics, No 10, The Publishing House of the Wroclav University of Economics*. Wroclav, 3(40).

[4] Vapnik V 1998 Statistical Learning Theory. *NY: John Wiley*

[5] Avriel and Mordecai 2003 Nonlinear Programming: Analysis and Methods. — *Dover Publishing*

[6] Castleman K 1979 Digital linage Processing. *Prentice-Hall: Englewood Cliffs*. – 407 p.

[7] Gray, Lawrence F. Flanigan, Francis J. Kazdan, Jerry L., Frank and David H 1990 Calculus two: linear and nonlinear functions, *Berlin: Springer-Verlag*, p 375

[8] Bartlett P and Shawe-Taylor J 1998 Generalization performance of support vector machines and other pattern classifiers // *Advances in Kernel Methods._ MIT Press, Cambridge*, USA. (http://citeseer.ist.psu.edu/bartlett98generalization.html)

Acknowledgments

I Express my gratitude to my supervisor, doctor of technical Sciences, prof. Igor Goroshko (Management Academy of the Ministry of the Interior of the Russian Federation), my teacher Ph. D., associate Professor. Korlyakova M. O. (Bauman Moscow State Technical University) and my father's Ph. D., Assoc. Ermolenko V. A. (Bauman Moscow State Technical University) for his invaluable contribution to the development of my scientific potential.